

TRAITE DE COOPERATION EN MATIERE DE BREVETS

PCT

NOTIFICATION D'ELECTION

(règle 61.2 du PCT)

Expéditeur: le BUREAU INTERNATIONAL

Destinataire:

United States Patent and Trademark
Office
(Box PCT)
Crystal Plaza 2
Washington, DC 20231
ÉTATS-UNIS D'AMÉRIQUE

en sa qualité d'office élu

Date d'expédition (jour/mois/année) 03 mars 1999 (03.03.99)	
Demande internationale no PCT/FR98/01460	Référence du dossier du déposant ou du mandataire MD/B05B2867
Date du dépôt international (jour/mois/année) 07 juillet 1998 (07.07.98)	Date de priorité (jour/mois/année) 07 juillet 1997 (07.07.97)
Déposant PARAHNOS-BACCALA, Glaucia etc	

1. L'office désigné est avisé de son élection qui a été faite:



dans la demande d'examen préliminaire international présentée à l'administration chargée de l'examen préliminaire international le:

01 février 1999 (01.02.99)



dans une déclaration visant une élection ultérieure déposée auprès du Bureau international le:

2. L'élection



a été faite



n'a pas été faite

avant l'expiration d'un délai de 19 mois à compter de la date de priorité ou, lorsque la règle 32 s'applique, dans le délai visé à la règle 32.2b).

Bureau international de l'OMPI 34, chemin des Colombettes 1211 Genève 20, Suisse no de télécopieur: (41-22) 740.14.35	Fonctionnaire autorisé Maria Victoria CORTIELLO no de téléphone: (41-22) 338.83.38
--	--



DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIÈRE DE BREVETS (PCT)

(51) Classification internationale des brevets ⁶ : C12N 15/48, C12Q 1/70, C07K 14/15, A61K 31/70	A3	(11) Numéro de publication internationale: WO 99/02666 (43) Date de publication internationale: 21 janvier 1999 (21.01.99)
(21) Numéro de la demande internationale: PCT/FR98/01460 (22) Date de dépôt international: 7 juillet 1998 (07.07.98) (30) Données relatives à la priorité: 97/08816 7 juillet 1997 (07.07.97) FR (71) Déposant (pour tous les Etats désignés sauf US): BIO MERIEUX [FR/FR]; Chemin de l'Orme, F-69280 Marcy l'Etoile (FR). (72) Inventeurs; et (75) Inventeurs/Déposants (US seulement): PARAH- NOS-BACCALA, Glauca [FR/FR]; 75, cours Duguesclin, F-69003 Lyon (FR). KOMURIAN-PRADEL, Florence [FR/FR]; 114, chemin du Pavillon, F-69250 Poleymieux au Mont d'Or (FR). BEDIN, Frédéric [FR/FR]; 6, rue Gaspard André, F-69002 Lyon (FR). SODOYER, Mireille [FR/FR]; 5, rue du Brûlet, F-69110 Sainte Foy les Lyon (FR). OTT, Catherine [FR/FR]; 103, avenue Berthelot, F-69007 Lyon (FR). MALLET, François [FR/FR]; 84, rue Anatole France, F-69100 Villeurbanne (FR). PERRON, Hervé [FR/FR]; 15, rue de Boyer, F-69005 Lyon (FR). MANDRAND, Bernard [FR/FR]; 21, rue de la Doua, F-69100 Villeurbanne (FR).		(74) Mandataire: CABINET GERMAIN & MAUREAU; Boîte postale 6153, F-69466 Lyon Cedex 06 (FR). (81) Etats désignés: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, brevet ARIPO (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), brevet eurasién (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Publiée <i>Avec rapport de recherche internationale.</i> <i>Avant l'expiration du délai prévu pour la modification des</i> <i>revendications, sera republiée si de telles modifications sont</i> <i>reçues.</i> (88) Date de publication du rapport de recherche internationale: 15 avril 1999 (15.04.99)
(54) Title: RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR ASSOCIATED WITH MUL- TIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES (54) Titre: MATERIEL NUCLEIQUE RETROVIRAL ET FRAGMENTS NUCLEOTIDIQUES NOTAMMENT ASSOCIES A LA SCLE- ROSE EN PLAQUES ET/OU LA POLYARTHRITE RHUMATOIDE, A DES FINS DE DIAGNOSTIC, PROPHYLAC- TIQUES ET THERAPEUTIQUES (57) Abstract <p>The invention concerns a nucleic material, in isolated or purified state, and a nucleotide fragment comprising a nucleotide sequence selected from the group consisting in (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142, (ii) the complementary sequences of sequences (i); and (iii) the sequences equivalent to sequences (ii) and (iii), in particular the sequence having for every series of 100 contiguous monomers, at least 50 %, preferably 70 % homology with sequences (i) and (ii) respectively. The invention also concerns their uses for detecting a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis.</p> (57) Abrégé <p>Matériel nucléique, à l'état isolé ou purifié, et fragment nucléotidique, comprenant une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 et SEQ ID NO: 142, (ii) les séquences complémentaires des séquences (i); et (iii) les séquences équivalentes aux séquences (i) ou (ii), en particulier les séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins 70 % d'homologie avec respectivement les séquences (i) ou (ii), et utilisations pour détecter un rétrovirus associé à la sclérose en plaques et/ou la polyarthrite rhumatoïde.</p>		

UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AL	Albanie	ES	Espagne	LS	Lesotho	SI	Slovénie
AM	Arménie	FI	Finlande	LT	Lituanie	SK	Slovaquie
AT	Autriche	FR	France	LU	Luxembourg	SN	Sénégal
AU	Australie	GA	Gabon	LV	Lettonie	SZ	Swaziland
AZ	Azerbaïdjan	GB	Royaume-Uni	MC	Monaco	TD	Tchad
BA	Bosnie-Herzégovine	GE	Géorgie	MD	République de Moldova	TG	Togo
BB	Barbade	GH	Ghana	MG	Madagascar	TJ	Tadjikistan
BE	Belgique	GN	Guinée	MK	Ex-République yougoslave de Macédoine	TM	Turkménistan
BF	Burkina Faso	GR	Grèce	ML	Mali	TR	Turquie
BG	Bulgarie	HU	Hongrie	MN	Mongolie	TT	Trinité-et-Tobago
BJ	Bénin	IE	Irlande	MR	Mauritanie	UA	Ukraine
BR	Brésil	IL	Israël	MW	Malawi	UG	Ouganda
BY	Bélarus	IS	Islande	MX	Mexique	US	Etats-Unis d'Amérique
CA	Canada	IT	Italie	NE	Niger	UZ	Ouzbékistan
CF	République centrafricaine	JP	Japon	NL	Pays-Bas	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norvège	YU	Yougoslavie
CH	Suisse	KG	Kirghizistan	NZ	Nouvelle-Zélande	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	République populaire démocratique de Corée	PL	Pologne		
CM	Cameroun	KR	République de Corée	PT	Portugal		
CN	Chine	KZ	Kazakstan	RO	Roumanie		
CU	Cuba	LC	Sainte-Lucie	RU	Fédération de Russie		
CZ	République tchèque	LI	Liechtenstein	SD	Soudan		
DE	Allemagne	LK	Sri Lanka	SE	Suède		
DK	Danemark	LR	Libéria	SG	Singapour		
EE	Estonie						

RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale No
P. /FR 98/01460

A. CLASSEMENT DE L'OBJET DE LA DEMANDE

CIB 6 C12N15/48 C12Q1/70 C07K14/15 A61K31/70

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)

CIB 6 C12N C07K C12Q A61K

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
P,X	WO 98 23755 A (BIO MERIEUX) 4 juin 1998 voir le document en entier --- -/--	1-26

☒ Voir la suite du cadre C pour la fin de la liste des documents

☒ Les documents de familles de brevets sont indiqués en annexe

* Catégories spéciales de documents cités:

- "A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- "E" document antérieur, mais publié à la date de dépôt international ou après cette date
- "L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- "O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- "P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

- "T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention
- "X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément
- "Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier
- "&" document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

29 janvier 1999

Date d'expédition du présent rapport de recherche internationale

17.02.99

Nom et adresse postale de l'administration chargée de la recherche internationale

Office Européen des Brevets, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Cupido, M

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	<p>DATABASE EMBLHTG SEQ ID NO: HSAC64 Clone humain BAC RG083M05 de 7q21-7q22, 14 février 1997 PAULEY A: "The sequence of H. sapiens BAC clone RG083M05" XP002090413 Comparez des nucléotides 31070-32453 de RG083M05 avec nucléotides 120-1511 de SEQ ID NO:130, nucléotides 35881-37340 de RG085M05 avec nucléotides 1-1462 de SEQ ID NO:117, nucléotides 37331-37876 de RG083M05 avec nucléotides 70-620 de SEQ ID NO:114 et nucléotides 37333-38260 de RG083M05 avec nucléotides 330-1260 de SEQ ID NO:120</p> <p>----</p>	1-18,22
X	<p>DATABASE EMBLHUM2 SEQ ID HSU85196 Homo sapiens BAC378, séquence complète, 27 mai 1997 BOYSEN C ET AL.: "Analysis of the 1.1-Mb human alpha/delta T-cell receptor locus with bacterial artificial chromosome clones" XP002090414 Comparez nucléotides 7150-7510 de HSU85196 avec nucléotides 1-361 de SEQ ID NO:141 et SEQ ID NO:142</p> <p>----</p>	1,11, 13-15, 18-20, 22-26
X	<p>DATABASE EMBLEMEST11 SEQ ID HSZZ32436 AC AA327384 EST30710 Colon I Homo sapiens cDNA 5', 18 avril 1997 ADAMS M D ET AL.: "Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence" XP002059257 & NATURE., vol. 377supp, 28 septembre 1995, pages 3-174, LONDON GB</p> <p>----</p>	1,2, 14-20
X	<p>FR 2 737 500 A (BIO MERIEUX S.A.) 7 février 1997 cité dans la demande voir le document en entier</p> <p>----</p>	1-3,5, 13-20, 22-26
X	<p>WO 95 21256 A (BIO MERIEUX S.A.) 10 août 1995 cité dans la demande voir le document en entier</p> <p>----</p>	1-3,5, 13-20, 22-26
	<p>----</p> <p>-/--</p>	

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	G. LA MANTIA ET AL.: "Identification and characterization of novel human endogenous retroviral sequences preferentially expressed in undifferentiated embryonal carcinoma cells" NUCLEIC ACIDS RESEARCH, vol. 19, no. 7, 1991, pages 1513-1520, XP002059255 OXFORD GB voir le document en entier ---	1-3,5, 13-17, 22-24
P,X	H. PERRON ET AL.: "Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 94, no. 14, 8 juillet 1997, pages 7583-7588, XP002059256 WASHINGTON US voir le document en entier ---	1-3,5, 13-20, 22-26
A	WO 94 28138 A (UNIVERSITY COLLEGE LONDON) 8 décembre 1994 voir le document en entier ---	1-3,5, 13-20, 22-26
A	WO 93 07259 A (SCLEROSE-FORENINGEN) 15 avril 1993 voir le document en entier -----	1-3,5, 13-20, 22-26

Cadre I Observations - lorsqu'il a été estimé que certaines revendications ne pouvaient pas faire l'objet d'une recherche (suite du point 1 de la première feuille)

Conformément à l'article 17.2)a), certaines revendications n'ont pas fait l'objet d'une recherche pour les motifs suivants:

1. ☐ Les revendications n^{os} se rapportent à un objet à l'égard duquel l'administration n'est pas tenue de procéder à la recherche, à savoir:
2. ☐ Les revendications n^{os} se rapportent à des parties de la demande internationale qui ne remplissent pas suffisamment les conditions prescrites pour qu'une recherche significative puisse être effectuée, en particulier:
3. ☐ Les revendications n^{os} sont des revendications dépendantes et ne sont pas rédigées conformément aux dispositions de la deuxième et de la troisième phrases de la règle 6.4.a).

Cadre II Observations - lorsqu'il y a absence d'unité de l'invention (suite du point 2 de la première feuille)

L'administration chargée de la recherche internationale a trouvé plusieurs inventions dans la demande internationale, à savoir:

voir feuille supplémentaire

1. ☒ Comme toutes les taxes additionnelles ont été payées dans les délais par le déposant, le présent rapport de recherche internationale porte sur toutes les revendications pouvant faire l'objet d'une recherche.
2. ☐ Comme toutes les recherches portant sur les revendications qui s'y prêtaient ont pu être effectuées sans effort particulier justifiant une taxe additionnelle, l'administration n'a sollicité le paiement d'aucune taxe de cette nature.
3. ☐ Comme une partie seulement des taxes additionnelles demandées a été payée dans les délais par le déposant, le présent rapport de recherche internationale ne porte que sur les revendications pour lesquelles les taxes ont été payées, à savoir les revendications n^{os}
4. ☐ Aucune taxe additionnelle demandée n'a été payée dans les délais par le déposant. En conséquence, le présent rapport de recherche internationale ne porte que sur l'invention mentionnée en premier lieu dans les revendications; elle est couverte par les revendications n^{os}

Remarque quant à la réserve

- ☒ Les taxes additionnelles étaient accompagnées d'une réserve de la part du déposant.
- ☐ Le paiement des taxes additionnelles n'était assorti d'aucune réserve.

SUITE DES RENSEIGNEMENTS INDIQUES SUR PCT/ISA/ 210

L'administration chargée de la recherche internationale a trouvé plusieurs (groupes d') inventions dans la demande internationale, à savoir:

1. revendications: 5 (complètement), 1-3, 13-20, 22-26 (partiellement)

Matériel nucléaire comprenant une séquence SEQ ID NO:112, ou codant pour une peptide avec la séquence SEQ ID NO:113, peptides correspondants, compositions comprenant un fragment desdites séquences et leur utilisation dans un procédé pour détecter un rétrovirus associé à la sclérose en plaques.

2. revendications: 10 (complètement) 1, 2, 7-9, 11-26 (partiellement)

Comme pour l'invention 1 mais se rapportant à SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:120 et SEQ ID NO:121.

3. revendications: 1, 3, 13-15, 18, 19, 21-26 (toutes partiellement)

Comme pour l'invention 1 mais se rapportant à SEQ ID NO:124.

4. revendications: 1, 2, 4, 6, 13-19, 21-26 (toutes partiellement)

Comme pour l'invention 1 mais se rapportant à SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:135 et SEQ ID NO:137.

5. revendications: 1, 11, 13-15, 18-20, 22-26 (toutes partiellement)

Comme pour l'invention 1 mais se rapportant à SEQ ID NO:141 et SEQ ID NO:142.

RAPPORT DE RECHERCH INTERNATIONALE

Renseignements relatifs aux me s de familles de brevets

Demon. nationale No
P /FR 98/01460

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 9823755 A	04-06-1998	AUCUN	
FR 2737500 A	07-02-1997	AU 6823296 A BG 101355 A BR 9606566 A CA 2201282 A CZ 9701357 A EP 0789077 A WO 9706260 A NO 971493 A PL 319512 A	05-03-1997 30-12-1997 30-12-1997 20-02-1997 17-06-1998 13-08-1997 20-02-1997 03-06-1997 18-08-1997
WO 9521256 A	10-08-1995	FR 2715936 A FR 2715938 A FR 2715939 A FR 2715937 A FR 2727428 A FR 2728585 A AU 1711495 A CA 2141907 A DE 674004 T EP 0674004 A FI 954699 A JP 8511170 T NO 953925 A NZ 279855 A US 5800980 A	11-08-1995 11-08-1995 11-08-1995 11-08-1995 31-05-1996 28-06-1996 21-08-1995 05-08-1995 19-09-1996 27-09-1995 03-10-1995 26-11-1996 04-12-1995 27-05-1998 01-09-1998
WO 9428138 A	08-12-1994	AU 676568 B AU 6760094 A CA 2163641 A EP 0700441 A JP 8511936 T	13-03-1997 20-12-1994 08-12-1994 13-03-1996 17-12-1996
WO 9307259 A	15-04-1993	AU 664049 B AU 2770992 A CA 2121030 A EP 0609305 A	02-11-1995 03-05-1993 15-04-1993 10-08-1994

PCT**REQUEST**

The undersigned requests that the present international application be processed according to the Patent Cooperation Treaty.

International Application No.

International Filing Date

Name of receiving Office and "PCT International Application"

Applicant's or agent's file reference
(if desired) (12 characters maximum) MD/B05B2867

Box No. I TITLE OF INVENTION

RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES

Box No. II APPLICANT

Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)

BIO MERIEUX
Chemin de l'Orme
69280 MARCY L'ETOILE
FRANCE

☐ This person is also inventor.

Telephone No.

Facsimile No.

Teleprinter No.

State (that is, country) of nationality:

FRANCE

State (that is, country) of residence:

FRANCE

This person is applicant
for the purposes of:

☐all designated
States☒all designated States except the
United States of America☐the United States
of America only☐the States indicated in
the Supplemental Box**Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)**

Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)

PARAHNOS-BACCALA Glauca
75 Cours Duguesclin
69003 LYON
FRANCE

This person is:

☐ applicant only

☒ applicant and inventor

☐ inventor only (If this check-box
is marked, do not fill in below.)

State (that is, country) of nationality:

FRANCE

State (that is, country) of residence:

FRANCE

This person is applicant
for the purposes of:

☐all designated
States☐all designated States except
the United States of America☒the United States
of America only☐the States indicated in
the Supplemental Box

☒ Further applicants and/or (further) inventors are indicated on a continuation sheet.

Box No. IV AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE

The person identified below is hereby/has been appointed to act on behalf
of the applicant(s) before the competent International Authorities as:

☒

agent

☐

common representative

Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)

CABINET GERMAIN & MAUREAU
B.P. 6153
69466 LYON CEDEX 06
FRANCE

Telephone No.

04 72 69 84 30

Facsimile No.

04 72 69 84 31

Teleprinter No.

☐ Address for correspondence: Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.

Continuation of Box No. III		FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)	
<i>If none of the following sub-boxes is used, this sheet should not be included in the request.</i>			
Name and address: <i>(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)</i> KOMURIAN-PRADEL Florence 114 Chemin du Pavillon 69250 POLEYMIEUX AU MONT D'OR FRANCE		This person is: <input type="checkbox"/> applicant only <input checked="" type="checkbox"/> applicant and inventor <input type="checkbox"/> inventor only <i>(If this check-box is marked, do not fill in below.)</i>	
State <i>(that is, country)</i> of nationality: FRANCE		State <i>(that is, country)</i> of residence: FRANCE	
This person is applicant for the purposes of: <input type="checkbox"/> all designated States		<input type="checkbox"/> all designated States except the United States of America <input checked="" type="checkbox"/> the United States of America only <input type="checkbox"/> the States indicated in the Supplemental Box	
Name and address: <i>(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)</i> BEDIN Frederic 6 Rue Gaspard Andre 69002 LYON FRANCE		This person is: <input type="checkbox"/> applicant only <input checked="" type="checkbox"/> applicant and inventor <input type="checkbox"/> inventor only <i>(If this check-box is marked, do not fill in below.)</i>	
State <i>(that is, country)</i> of nationality: FRANCE		State <i>(that is, country)</i> of residence: FRANCE	
This person is applicant for the purposes of: <input type="checkbox"/> all designated States		<input type="checkbox"/> all designated States except the United States of America <input checked="" type="checkbox"/> the United States of America only <input type="checkbox"/> the States indicated in the Supplemental Box	
Name and address: <i>(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)</i> SODOYER Mireille 5 rue du Brulet 69110 SAINTE FOY LES LYON FRANCE		This person is: <input type="checkbox"/> applicant only <input checked="" type="checkbox"/> applicant and inventor <input type="checkbox"/> inventor only <i>(If this check-box is marked, do not fill in below.)</i>	
State <i>(that is, country)</i> of nationality: FRANCE		State <i>(that is, country)</i> of residence: FRANCE	
This person is applicant for the purposes of: <input type="checkbox"/> all designated States		<input type="checkbox"/> all designated States except the United States of America <input checked="" type="checkbox"/> the United States of America only <input type="checkbox"/> the States indicated in the Supplemental Box	
Name and address: <i>(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)</i> OTT Catherine 103 Avenue Berthelot 69007 LYON FRANCE		This person is: <input type="checkbox"/> applicant only <input checked="" type="checkbox"/> applicant and inventor <input type="checkbox"/> inventor only <i>(If this check-box is marked, do not fill in below.)</i>	
State <i>(that is, country)</i> of nationality: FRANCE		State <i>(that is, country)</i> of residence: FRANCE	
This person is applicant for the purposes of: <input type="checkbox"/> all designated States		<input type="checkbox"/> all designated States except the United States of America <input checked="" type="checkbox"/> the United States of America only <input type="checkbox"/> the States indicated in the Supplemental Box	
<input checked="" type="checkbox"/> Further applicants and/or (further) inventors are indicated on another continuation sheet.			

Continuation of Box No. III		FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)	
<i>If none of the following sub-boxes is used, this sheet should not be included in the request.</i>			
Name and address: <i>(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)</i> MALLET Francois 84 rue Anatole France 69100 VILLEURBANNE FRANCE		This person is: <input type="checkbox"/> applicant only <input checked="" type="checkbox"/> applicant and inventor <input type="checkbox"/> inventor only <i>(If this check-box is marked, do not fill in below.)</i>	
State <i>(that is, country)</i> of nationality: FRANCE		State <i>(that is, country)</i> of residence: FRANCE	
This person is applicant for the purposes of: <input type="checkbox"/> all designated States <input type="checkbox"/> all designated States except the United States of America <input checked="" type="checkbox"/> the United States of America only <input type="checkbox"/> the States indicated in the Supplemental Box			
Name and address: <i>(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)</i> PERRON Herve 15 rue de Boyer 69005 LYON FRANCE		This person is: <input type="checkbox"/> applicant only <input checked="" type="checkbox"/> applicant and inventor <input type="checkbox"/> inventor only <i>(If this check-box is marked, do not fill in below.)</i>	
State <i>(that is, country)</i> of nationality: FRANCE		State <i>(that is, country)</i> of residence: FRANCE	
This person is applicant for the purposes of: <input type="checkbox"/> all designated States <input type="checkbox"/> all designated States except the United States of America <input checked="" type="checkbox"/> the United States of America only <input type="checkbox"/> the States indicated in the Supplemental Box			
Name and address: <i>(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)</i> MANDRAND Bernard 21 rue de la Doua 69100 VILLEURBANNE FRANCE		This person is: <input type="checkbox"/> applicant only <input checked="" type="checkbox"/> applicant and inventor <input type="checkbox"/> inventor only <i>(If this check-box is marked, do not fill in below.)</i>	
State <i>(that is, country)</i> of nationality: FRANCE		State <i>(that is, country)</i> of residence: FRANCE	
This person is applicant for the purposes of: <input type="checkbox"/> all designated States <input type="checkbox"/> all designated States except the United States of America <input checked="" type="checkbox"/> the United States of America only <input type="checkbox"/> the States indicated in the Supplemental Box			
Name and address: <i>(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)</i> 		This person is: <input type="checkbox"/> applicant only <input type="checkbox"/> applicant and inventor <input type="checkbox"/> inventor only <i>(If this check-box is marked, do not fill in below.)</i>	
State <i>(that is, country)</i> of nationality:		State <i>(that is, country)</i> of residence:	
This person is applicant for the purposes of: <input type="checkbox"/> all designated States <input type="checkbox"/> all designated States except the United States of America <input type="checkbox"/> the United States of America only <input type="checkbox"/> the States indicated in the Supplemental Box			
<input type="checkbox"/> Further applicants and/or (further) inventors are indicated on another continuation sheet.			

Box No. V DESIGNATION OF STATES

The following designations are hereby made under Rule 4.9(a) (mark the applicable check-boxes; at least one must be marked):

Regional Patent

- ☒ **AP ARIPO Patent:** GH Ghana, GM Gambia, KE Kenya, LS Lesotho, MW Malawi, SD Sudan, SZ Swaziland, UG Uganda, ZW Zimbabwe, and any other State which is a Contracting State of the Harare Protocol and of the PCT
- ☒ **EA Eurasian Patent:** AM Armenia, AZ Azerbaijan, BY Belarus, KG Kyrgyzstan, KZ Kazakhstan, MD Republic of Moldova, RU Russian Federation, TJ Tajikistan, TM Turkmenistan, and any other State which is a Contracting State of the Eurasian Patent Convention and of the PCT
- ☒ **EP European Patent:** AT Austria, BE Belgium, CH and LI Switzerland and Liechtenstein, CY Cyprus, DE Germany, DK Denmark, ES Spain, FI Finland, FR France, GB United Kingdom, GR Greece, IE Ireland, IT Italy, LU Luxembourg, MC Monaco, NL Netherlands, PT Portugal, SE Sweden, and any other State which is a Contracting State of the European Patent Convention and of the PCT
- ☒ **OA OAPI Patent:** BF Burkina Faso, BJ Benin, CF Central African Republic, CG Congo, CI Côte d'Ivoire, CM Cameroon, GA Gabon, GN Guinea, ML Mali, MR Mauritania, NE Niger, SN Senegal, TD Chad, TG Togo, and any other State which is a member State of OAPI and a Contracting State of the PCT (if other kind of protection or treatment desired, specify on dotted line)

National Patent (if other kind of protection or treatment desired, specify on dotted line):

- | | |
|---|---|
| <input checked="" type="checkbox"/> AL Albania | <input checked="" type="checkbox"/> LS Lesotho |
| <input checked="" type="checkbox"/> AM Armenia | <input checked="" type="checkbox"/> LT Lithuania |
| <input checked="" type="checkbox"/> AT Austria | <input checked="" type="checkbox"/> LU Luxembourg |
| <input checked="" type="checkbox"/> AU Australia | <input checked="" type="checkbox"/> LV Latvia |
| <input checked="" type="checkbox"/> AZ Azerbaijan | <input checked="" type="checkbox"/> MD Republic of Moldova |
| <input checked="" type="checkbox"/> BA Bosnia and Herzegovina | <input checked="" type="checkbox"/> MG Madagascar |
| <input checked="" type="checkbox"/> BB Barbados | <input checked="" type="checkbox"/> MK The former Yugoslav Republic of Macedonia |
| <input checked="" type="checkbox"/> BG Bulgaria | |
| <input checked="" type="checkbox"/> BR Brazil | <input checked="" type="checkbox"/> MN Mongolia |
| <input checked="" type="checkbox"/> BY Belarus | <input checked="" type="checkbox"/> MW Malawi |
| <input checked="" type="checkbox"/> CA Canada | <input checked="" type="checkbox"/> MX Mexico |
| <input checked="" type="checkbox"/> CH and LI Switzerland and Liechtenstein | <input checked="" type="checkbox"/> NO Norway |
| <input checked="" type="checkbox"/> CN China | <input checked="" type="checkbox"/> NZ New Zealand |
| <input checked="" type="checkbox"/> CU Cuba | <input checked="" type="checkbox"/> PL Poland |
| <input checked="" type="checkbox"/> CZ Czech Republic | <input checked="" type="checkbox"/> PT Portugal |
| <input checked="" type="checkbox"/> DE Germany | <input checked="" type="checkbox"/> RO Romania |
| <input checked="" type="checkbox"/> DK Denmark | <input checked="" type="checkbox"/> RU Russian Federation |
| <input checked="" type="checkbox"/> EE Estonia | <input checked="" type="checkbox"/> SD Sudan |
| <input checked="" type="checkbox"/> ES Spain | <input checked="" type="checkbox"/> SE Sweden |
| <input checked="" type="checkbox"/> FI Finland | <input checked="" type="checkbox"/> SG Singapore |
| <input checked="" type="checkbox"/> GB United Kingdom | <input checked="" type="checkbox"/> SI Slovenia |
| <input checked="" type="checkbox"/> GE Georgia | <input checked="" type="checkbox"/> SK Slovakia |
| <input checked="" type="checkbox"/> GH Ghana | <input checked="" type="checkbox"/> SL Sierra Leone |
| <input checked="" type="checkbox"/> GM Gambia | <input checked="" type="checkbox"/> TJ Tajikistan |
| <input checked="" type="checkbox"/> HR Croatia | <input checked="" type="checkbox"/> TM Turkmenistan |
| <input checked="" type="checkbox"/> HU Hungary | <input checked="" type="checkbox"/> TR Turkey |
| <input checked="" type="checkbox"/> ID Indonesia | <input checked="" type="checkbox"/> TT Trinidad and Tobago |
| <input checked="" type="checkbox"/> IL Israel | <input checked="" type="checkbox"/> UA Ukraine |
| <input checked="" type="checkbox"/> IS Iceland | <input checked="" type="checkbox"/> UG Uganda |
| <input checked="" type="checkbox"/> JP Japan | <input checked="" type="checkbox"/> US United States of America |
| <input checked="" type="checkbox"/> KE Kenya | |
| <input checked="" type="checkbox"/> KG Kyrgyzstan | <input checked="" type="checkbox"/> UZ Uzbekistan |
| <input checked="" type="checkbox"/> KP Democratic People's Republic of Korea | <input checked="" type="checkbox"/> VN Viet Nam |
| | <input checked="" type="checkbox"/> YU Yugoslavia |
| <input checked="" type="checkbox"/> KR Republic of Korea | <input checked="" type="checkbox"/> ZW Zimbabwe |
| <input checked="" type="checkbox"/> KZ Kazakhstan | |
| <input checked="" type="checkbox"/> LC Saint Lucia | |
| <input checked="" type="checkbox"/> LK Sri Lanka | |
| <input checked="" type="checkbox"/> LR Liberia | |

Check-boxes reserved for designating States (for the purposes of a national patent) which have become party to the PCT after issuance of this sheet:

☐

☐

Precautionary Designation Statement: In addition to the designations made above, the applicant also makes under Rule 4.9(b) all other designations which would be permitted under the PCT except the designation(s) indicated in the Supplemental Box as being excluded from the scope of this statement. The applicant declares that those additional designations are subject to confirmation and that any designation which is not confirmed before the expiration of 15 months from the priority date is to be regarded as withdrawn by the applicant at the expiration of that time limit. (Confirmation of a designation consists of the filing of a notice specifying that designation and the payment of the designation and confirmation fees. Confirmation must reach the receiving Office within the 15-month time limit.)

Box No. VI PRIORITY CLAIM		<input type="checkbox"/> Further priority claims are indicated in the Supplemental Box.		
Filing date of earlier application (day/month/year)	Number of earlier application	Where earlier application is:		
		national application: country	regional application:* regional Office	international application: receiving Office
item (1) 7 July 1997	97 08816	FRANCE		
item (2)				
item (3)				

☒ The receiving Office is requested to prepare and transmit to the International Bureau a certified copy of the earlier application(s) (only if the earlier application was filed with the Office which for the purposes of the present international application is the receiving Office) identified above as item(s): (1)

* Where the earlier application is an ARIPO application, it is mandatory to indicate in the Supplemental Box at least one country party to the Paris Convention for the Protection of Industrial Property for which that earlier application was filed (Rule 4.10(b)(ii)). See Supplemental Box.

Box No. VII INTERNATIONAL SEARCHING AUTHORITY

Choice of International Searching Authority (ISA) <i>(if two or more International Searching Authorities are competent to carry out the international search, indicate the Authority chosen; the two-letter code may be used):</i> ISA /	Request to use results of earlier search: reference to that search (if an earlier search has been carried out by or requested from the International Searching Authority): Date (day/month/year) Number Country (or regional Office)
---	--

Box No. VIII CHECK LIST; LANGUAGE OF FILING

This international application contains the following number of sheets: request :5 description (excluding sequence listing part) :38 claims :5 abstract :1 drawings :32 sequence listing part of description :20 Total number of sheets :101	This international application is accompanied by the item(s) marked below: 1. <input type="checkbox"/> fee calculation sheet 2. <input type="checkbox"/> separate signed power of attorney 3. <input checked="" type="checkbox"/> copy of general power of attorney; reference number, if any: 4. <input type="checkbox"/> statement explaining lack of signature 5. <input type="checkbox"/> priority document(s) identified in Box No. VI as item(s): 6. <input type="checkbox"/> translation of international application into (language): 7. <input type="checkbox"/> separate indications concerning deposited microorganism or other biological material 8. <input type="checkbox"/> nucleotide and/or amino acid sequence listing in computer readable form 9. <input type="checkbox"/> other (specify):
--	--

Figure of the drawings which should accompany the abstract:	Language of filing of the international application: FRENCH
--	--

Box No. IX SIGNATURE OF APPLICANT OR AGENT

Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such capacity is not obvious from reading the request).

CABINET GERMAIN & MAUREAU

Dominique GUERRE
CPI 321104

Lyon, 7 July 1998

For receiving Office use only

1. Date of actual receipt of the purported international application: 3. Corrected date of actual receipt due to later but timely received papers or drawings completing the purported international application: 4. Date of timely receipt of the required corrections under PCT Article 11(2): 5. International Searching Authority (if two or more are competent): ISA /	2. Drawings: <input type="checkbox"/> received: <input type="checkbox"/> not received: 6. <input type="checkbox"/> Transmittal of search copy delayed until search fee is paid
--	---

For International Bureau use only

Date of receipt of the record copy by the International Bureau:	
---	--

PCT

RAPPORT DE RECHERCHE INTERNATIONALE

(article 18 et règles 43 et 44 du PCT)

Référence du dossier du déposant ou du mandataire MD/B05B2867	POUR SUITE A DONNER voir la notification de transmission du rapport de recherche internationale (formulaire PCT/ISA/220) et, le cas échéant, le point 5 ci-après	
Demande internationale n° PCT/ FR 98/ 01460	Date du dépôt international (jour/mois/année) 07/07/1998	(Date de priorité (la plus ancienne) (jour/mois/année) 07/07/1997
Déposant BIO MERIEUX et al.		

Le présent rapport de recherche internationale, établi par l'administration chargée de la recherche internationale, est transmis au déposant conformément à l'article 18. Une copie en est transmise au Bureau international.

Ce rapport de recherche internationale comprend 6 feuilles.

☒ Il est aussi accompagné d'une copie de chaque document relatif à l'état de la technique qui y est cité.

1. ☐ Il a été estimé que certaines revendications ne pouvaient pas faire l'objet d'une recherche (voir le cadre I).

2. ☒ Il y a absence d'unité de l'invention (voir le cadre II).

3. ☒ La demande internationale contient la divulgation d'un listage de séquence de nucléotides ou d'acides aminés et la recherche internationale a été effectuée sur la base du listage de séquence

☒ déposé avec la demande internationale

☐ fourni par le déposant séparément de la demande internationale

☐ sans être accompagnée d'une déclaration selon laquelle il n'inclut pas d'éléments allant au-delà de la divulgation faite dans la demande internationale telle qu'elle a été déposée.

☐ transcrit par l'administration

4. En ce qui concerne le titre, ☒ le texte est approuvé tel qu'il a été remis par le déposant.

☐ Le texte a été établi par l'administration et a la teneur suivante:

5. En ce qui concerne l'abrégé,

☒ le texte est approuvé tel qu'il a été remis par le déposant

☐ le texte (reproduit dans le cadre III) a été établi par l'administration conformément à la règle 38.2b). Le déposant peut présenter des observations à l'administration dans un délai d'un mois à compter de la date d'expédition du présent rapport de recherche internationale.

6. La figure des dessins à publier avec l'abrégé est la suivante:

Figure n° ☐ suggérée par le déposant.

☐ parce que le déposant n'a pas suggéré de figure.

☐ parce que cette figure caractérise mieux l'invention.

☒ Aucune des figures n'est à publier.

Cadre I Observations - lorsqu'il a été estimé que certaines revendications ne pouvaient pas faire l'objet d'une recherche (suite du point 1 de la première feuille)

Conformément à l'article 17.2)a), certaines revendications n'ont pas fait l'objet d'une recherche pour les motifs suivants:

1. ☐ Les revendications n^{os} se rapportent à un objet à l'égard duquel l'administration n'est pas tenue de procéder à la recherche, à savoir:
2. ☐ Les revendications n^{os} se rapportent à des parties de la demande internationale qui ne remplissent pas suffisamment les conditions prescrites pour qu'une recherche significative puisse être effectuée, en particulier:
3. ☐ Les revendications n^{os} sont des revendications dépendantes et ne sont pas rédigées conformément aux dispositions de la deuxième et de la troisième phrases de la règle 6.4.a).

Cadre II Observations - lorsqu'il y a absence d'unité de l'invention (suite du point 2 de la première feuille)

L'administration chargée de la recherche internationale a trouvé plusieurs inventions dans la demande internationale, à savoir:

voir feuille supplémentaire

1. ☒ Comme toutes les taxes additionnelles ont été payées dans les délais par le déposant, le présent rapport de recherche internationale porte sur toutes les revendications pouvant faire l'objet d'une recherche.
2. ☐ Comme toutes les recherches portant sur les revendications qui s'y prêtaient ont pu être effectuées sans effort particulier justifiant une taxe additionnelle, l'administration n'a sollicité le paiement d'aucune taxe de cette nature.
3. ☐ Comme une partie seulement des taxes additionnelles demandées a été payée dans les délais par le déposant, le présent rapport de recherche internationale ne porte que sur les revendications pour lesquelles les taxes ont été payées, à savoir les revendications n^{os}
4. ☐ Aucune taxe additionnelle demandée n'a été payée dans les délais par le déposant. En conséquence, le présent rapport de recherche internationale ne porte que sur l'invention mentionnée en premier lieu dans les revendications; elle est couverte par les revendications n^{os}

Remarque quant à la réserve

☒ Les taxes additionnelles étaient accompagnées d'une réserve de la part du déposant.

☐ Le paiement des taxes additionnelles n'était assorti d'aucune réserve.

SUITE DES RENSEIGNEMENTS INDICUES SUR PCT/ISA/ 210

L'administration chargée de la recherche internationale a trouvé plusieurs (groupes d') inventions dans la demande internationale, à savoir:

1. revendications: 5 (complètement), 1-3, 13-20, 22-26 (partiellement)

Matériel nucléaire comprenant une séquence SEQ ID NO:112, ou codant pour une peptide avec la séquence SEQ ID NO:113, peptides correspondants, compositions comprenant un fragment desdites séquences et leur utilisation dans un procédé pour détecter un rétrovirus associé à la sclérose en plaques.

2. revendications: 10 (complètement) 1, 2, 7-9, 11-26 (partiellement)

Comme pour l'invention 1 mais se rapportant à SEQ ID NO:114, SEQ ID NO:115, SEQ ID NO:117, SEQ ID NO:118, SEQ ID NO:120 et SEQ ID NO:121.

3. revendications: 1, 3, 13-15, 18, 19, 21-26 (toutes partiellement)

Comme pour l'invention 1 mais se rapportant à SEQ ID NO:124.

4. revendications: 1, 2, 4, 6, 13-19, 21-26 (toutes partiellement)

Comme pour l'invention 1 mais se rapportant à SEQ ID NO:130, SEQ ID NO:131, SEQ ID NO:135 et SEQ ID NO:137.

5. revendications: 1, 11, 13-15, 18-20, 22-26 (toutes partiellement)

Comme pour l'invention 1 mais se rapportant à SEQ ID NO:141 et SEQ ID NO:142.

RAPPORT DE CHERCHE INTERNATIONALE

Demande internationale No

PCT/FR 98/01460

A. CLASSEMENT DE L'OBJET DE LA DEMANDE

CIB 6 C12N15/48 C12Q1/70 C07K14/15 A61K31/70

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)

CIB 6 C12N C07K C12Q A61K

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
P, X	<p>WO 98 23755 A (BIO MERIEUX) 4 juin 1998</p> <p>voir le document en entier</p> <p style="text-align: center;">---</p> <p style="text-align: center;">-/--</p>	1-26

☒ Voir la suite du cadre C pour la fin de la liste des documents

☒ Les documents de familles de brevets sont indiqués en annexe

* Catégories spéciales de documents cités:

"A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent

"E" document antérieur, mais publié à la date de dépôt international ou après cette date

"L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)

"O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens

"P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

"X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

"Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

"&" document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

29 janvier 1999

Date d'expédition du présent rapport de recherche internationale

17.02.99

Nom et adresse postale de l'administration chargée de la recherche internationale

Office Européen des Brevets, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Cupido, M

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	<p>DATABASE EMBLHTG SEQ ID NO: HSAC64 Clone humain BAC RG083M05 de 7q21-7q22, 14 février 1997 PAULEY A: "The sequence of H. sapiens BAC clone RG083M05" XP002090413 Comparez des nucléotides 31070-32453 de RG083M05 avec nucléotides 120-1511 de SEQ ID NO:130, nucléotides 35881-37340 de RG085M05 avec nucléotides 1-1462 de SEQ ID NO:117, nucléotides 37331-37876 de RG083M05 avec nucléotides 70-620 de SEQ ID NO:114 et nucléotides 37333-38260 de RG083M05 avec nucléotides 330-1260 de SEQ ID NO:120</p> <p>---</p>	1-18, 22
X	<p>DATABASE EMBLHUM2 SEQ ID HSU85196 Homo sapiens BAC378, séquence complète, 27 mai 1997 BOYSEN C ET AL.: "Analysis of the 1.1-Mb human alpha/delta T-cell receptor locus with bacterial artificial chromosome clones" XP002090414 Comparez nucléotides 7150-7510 de HSU85196 avec nucléotides 1-361 de SEQ ID NO:141 et SEQ ID NO:142</p> <p>---</p>	1, 11, 13-15, 18-20, 22-26
X	<p>DATABASE EMBLEMEST11 SEQ ID HSZZ32436 AC AA327384 EST30710 Colon I Homo sapiens cDNA 5', 18 avril 1997 ADAMS M D ET AL.: "Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence" XP002059257 & NATURE., vol. 377supp, 28 septembre 1995, pages 3-174, LONDON GB</p> <p>---</p>	1, 2, 14-20
X	<p>FR 2 737 500 A (BIO MERIEUX S.A.) 7 février 1997 cité dans la demande voir le document en entier</p> <p>---</p>	1-3, 5, 13-20, 22-26
X	<p>WO 95 21256 A (BIO MERIEUX S.A.) 10 août 1995 cité dans la demande voir le document en entier</p> <p>---</p>	1-3, 5, 13-20, 22-26
	<p>---</p> <p>-/--</p>	

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	G. LA MANTIA ET AL.: "Identification and characterization of novel human endogenous retroviral sequences preferentially expressed in undifferentiated embryonal carcinoma cells" NUCLEIC ACIDS RESEARCH, vol. 19, no. 7, 1991, pages 1513-1520, XP002059255 OXFORD GB voir le document en entier	1-3,5, 13-17, 22-24
P,X	H. PERRON ET AL.: "Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 94, no. 14, 8 juillet 1997, pages 7583-7588, XP002059256 WASHINGTON US voir le document en entier	1-3,5, 13-20, 22-26
A	WO 94 28138 A (UNIVERSITY COLLEGE LONDON) 8 décembre 1994 voir le document en entier	1-3,5, 13-20, 22-26
A	WO 93 07259 A (SCLEROSE-FORENINGEN) 15 avril 1993 voir le document en entier	1-3,5, 13-20, 22-26

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux familles de brevets

Recherche Internationale No

FR/FR 98/01460

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 9823755 A	04-06-1998	AUCUN	
FR 2737500 A	07-02-1997	AU 6823296 A	05-03-1997
		BG 101355 A	30-12-1997
		BR 9606566 A	30-12-1997
		CA 2201282 A	20-02-1997
		CZ 9701357 A	17-06-1998
		EP 0789077 A	13-08-1997
		WO 9706260 A	20-02-1997
		NO 971493 A	03-06-1997
		PL 319512 A	18-08-1997
WO 9521256 A	10-08-1995	FR 2715936 A	11-08-1995
		FR 2715938 A	11-08-1995
		FR 2715939 A	11-08-1995
		FR 2715937 A	11-08-1995
		FR 2727428 A	31-05-1996
		FR 2728585 A	28-06-1996
		AU 1711495 A	21-08-1995
		CA 2141907 A	05-08-1995
		DE 674004 T	19-09-1996
		EP 0674004 A	27-09-1995
		FI 954699 A	03-10-1995
		JP 8511170 T	26-11-1996
		NO 953925 A	04-12-1995
		NZ 279855 A	27-05-1998
		US 5800980 A	01-09-1998
WO 9428138 A	08-12-1994	AU 676568 B	13-03-1997
		AU 6760094 A	20-12-1994
		CA 2163641 A	08-12-1994
		EP 0700441 A	13-03-1996
		JP 8511936 T	17-12-1996
WO 9307259 A	15-04-1993	AU 664049 B	02-11-1995
		AU 2770992 A	03-05-1993
		CA 2121030 A	15-04-1993
		EP 0609305 A	10-08-1994

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application No :

U.S. National Serial No. :

Filed :

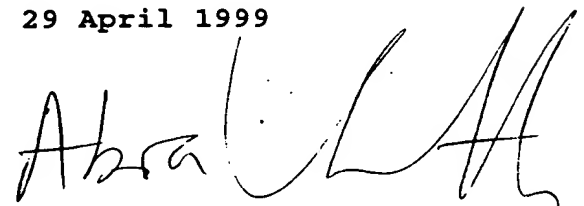
PCT International Application No. : PCT/FR98/01460

VERIFICATION OF A TRANSLATION

I, the below named translator, hereby declare that :
My name and post office address are as stated below ;
That I am knowledgeable in the French language in which the below identified international application was filed, and that, to the best of my knowledge and belief, the English translation of the international application No. PCT/FR98/01460 is a true and complete translation of the above identified international application as filed.

I hereby declare that all the statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that wilful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such wilful false statements may jeopardize the validity of the patent application issued thereon.

Date : 29 April 1999



Full name of the translator : Abraham SMITH

For and on behalf of RWS Group plc

Post Office Address :

Europa House, Marsham Way,
Gerrards Cross, Buckinghamshire,
England.

TRAITE DE COOPERATION EN MATIERE DE BREVETS

PCT

REC'D 11 NOV 1999

RAPPORT D'EXAMEN PRELIMINAIRE INTERNATIONAL

RECEIVED

DEC 27 1999

(article 36 et règle 70 du PCT)

Référence du dossier du déposant ou du mandataire MD/MK/B05B2867WO		voir la notification de transmission du rapport d'examen préliminaire international (formulaire PCT/IPEA/409)/2900	
Demande internationale n° PCT/FR98/01460		Date du dépôt international (jour/mois/année) 07/07/1998	Date de priorité (jour/mois/année) 07/07/1997
Classification internationale des brevets (CIB) ou à la fois classification nationale et CIB C12N15/00			
Déposant BIO MERIEUX et al.			

1. Le présent rapport d'examen préliminaire international, établi par l'administration chargée de l'examen préliminaire international, est transmis au déposant conformément à l'article 36.



2. Ce RAPPORT comprend 11 feuilles, y compris la présente feuille de couverture.

☒ Il est accompagné d'ANNEXES, c'est-à-dire de feuilles de la description, des revendications ou des dessins qui ont été modifiées et qui servent de base au présent rapport ou de feuilles contenant des rectifications faites auprès de l'administration chargée de l'examen préliminaire international (voir la règle 70.16 et l'instruction 607 des Instructions administratives du PCT).

Certaines annexes comprennent 6 feuilles.

3. Le présent rapport contient des indications relatives aux points suivants:

- I ☒ Base du rapport
- II ☒ Priorité
- III ☐ Absence de formulation d'opinion quant à la nouveauté, l'activité inventive et la possibilité d'application industrielle
- IV ☒ Absence d'unité de l'invention
- V ☒ Déclaration motivée selon l'article 35(2) quant à la nouveauté, l'activité inventive et la possibilité d'application industrielle; citations et explications à l'appui de cette déclaration
- VI ☒ Certains documents cités
- VII ☐ Irrégularités dans la demande internationale
- VIII ☒ Observations relatives à la demande internationale

Date de présentation de la demande d'examen préliminaire internationale 01/02/1999	Date d'achèvement du présent rapport 08. 11. 99
Nom et adresse postale de l'administration chargée de l'examen préliminaire international:  Office européen des brevets D-80298 Munich Tél. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Fonctionnaire autorisé BULCAO DE MELO ..., T N° de téléphone +49 89 2399 8972 

**RAPPORT D'EXAMEN
PRELIMINAIRE INTERNATIONAL**

Demande internationale n° PCT/FR98/01460

I. Base du rapport

1. Ce rapport a été rédigé sur la base des éléments ci-après (*les feuilles de remplacement qui ont été remises à l'office récepteur en réponse à une invitation faite conformément à l'article 14 sont considérées, dans le présent rapport, comme "initialement déposées" et ne sont pas jointes en annexe au rapport puisqu'elles ne contiennent pas de modifications.*) :

Description, pages:

1-38 version initiale

Revendications, N°:

1-28 reçue(s) le 18/10/1999 avec la lettre du 13/10/1999

Dessins, feuilles:

1/32-32/32 version initiale

2. Les modifications ont entraîné l'annulation :

- ☐ de la description, pages :
- ☐ des revendications, n°s :
- ☐ des dessins, feuilles :

3. ☐ Le présent rapport a été formulé abstraction faite (de certaines) des modifications, qui ont été considérées comme allant au-delà de l'exposé de l'invention tel qu'il a été déposé, comme il est indiqué ci-après (règle 70.2(c)) :

4. Observations complémentaires, le cas échéant :

II. Priorité

1. ☐ Le présent rapport a été formulée comme si aucune priorité n'avait été revendiquée, du fait que les documents suivants n'ont pas été remis dans le délai prescrit :
- ☐ copie de la demande antérieure dont la priorité a été revendiquée.
 - ☐ traduction de la demande antérieure dont la priorité a été revendiquée.
2. ☐ Le présent rapport a été formulée comme si aucune priorité n'avait été revendiquée, du fait que la

**RAPPORT D'EXAMEN
PRELIMINAIRE INTERNATIONAL**

Demande internationale n° PCT/FR98/01460

revendication de la priorité a été jugée non valable.

Pour les besoins du présent rapport, la date de dépôt international indiquée plus haut est donc considérée comme la date pertinente.

3. Observations complémentaires, le cas échéant :

voir feuille séparée

IV. Absence d'unité de l'invention

1. En réponse à l'invitation à limiter les revendications ou à payer des taxes additionnelles, le déposant a

- ☒ limité les revendications.
- ☐ payé des taxes additionnelles.
- ☐ payé des taxes additionnelles sous réserve.
- ☐ ni limité les revendications ni payé des taxes additionnelles.

2. ☐ L'administration chargée de l'examen préliminaire international estime qu'il n'est pas satisfait à l'exigence d'unité d'invention et décide, conformément à la règle 68.1, de ne pas inviter le déposant à limiter les revendications ou à payer des taxes additionnelles.

3. L'administration chargée de l'examen préliminaire international estime que, aux termes des règles 13.1, 13.2 et 13.3,

- ☐ il est satisfait à l'exigence d'unité de l'invention.
- ☒ il n'est pas satisfait à l'exigence d'unité de l'invention, et ce pour les raisons suivantes :

voir feuille séparée

4. En conséquence, les parties suivantes de la demande internationale ont fait l'objet d'un examen préliminaire international lors de la formulation du présent rapport :

- ☐ toutes les parties de la demande.
- ☒ les parties relatives aux revendications n°s 7-10 et 12 (en entier) et 1, 2, 11 et 13-28 (partiellement).

**RAPPORT D'EXAMEN
PRELIMINAIRE INTERNATIONAL**

Demande internationale n° PCT/FR98/01460

V. Déclaration motivée selon l'article 35(2) quant à la nouveauté, l'activité inventive et la possibilité d'application industrielle; citations et explications à l'appui de cette déclaration

1. Déclaration

Nouveauté	Oui : Revendications	7-10 et 12 (en entier) et 1, 2, 11 et 13-28 (partiellement)
	Non : Revendications	
Activité inventive	Oui : Revendications	
	Non : Revendications	7-10 et 12 (en entier) et 1, 2, 11 et 13-28 (partiellement)
Possibilité d'application industrielle	Oui : Revendications	7-10 et 12 (en entier) et 1, 2, 11 et 13-28 (partiellement)
	Non : Revendications	

2. Citations et explications

voir feuille séparée

VI. Certain documents cités

1. Certains documents publiés (règle 70.10)

et / ou

2. Divulgations non écrites (règle 70.9)

voir feuille séparée

VIII. Observations relatives à la demande internationale

Les observations suivantes sont faites au sujet de la clarté des revendications, de la description et des dessins et de la question de savoir si les revendications se fondent entièrement sur la description :

voir feuille séparée

SECTION I

1. Les **revendications modifiées 1-28**, déposées avec la lettre du 13.10.99, sont admissibles selon la **Règle 70.2(c) PCT**.

SECTION II

2. Le présent Rapport d'Examen Préliminaire International a été établi en ayant pour hypothèse que la **date de priorité 07.07.97** de la présente demande est valablement revendiquée. Le document Proceedings of the National Academy of Sciences of USA, Vol. 94, No. 14, 8 Juillet 1997, pages 7583-7588, cité comme document "P" dans le rapport de recherche n'a donc pas été considéré comme compris dans l'état de la technique selon le règlement d'exécution (**Règle 64 (1)-(3) PCT**).

SECTION IV

3. Absence d'Unité d'Invention (**Article 34(3) PCT et Règles 13 et 68 PCT**)
 - 3.1. L'administration chargée de l'examen préliminaire international estime que la présente demande internationale ne satisfait pas à l'exigence d'unité d'invention (**Article 34(3) PCT et Règles 13 et 68 PCT**).
 - 3.2. Une demande de brevet ne peut concerner qu'une invention ou une pluralité d'inventions liées entre elles de telle sorte qu'elles ne forment qu'un seul concept inventif général.

L'exigence d'unité d'invention (**Règle 13.1 PCT**) n'est observée que s'il existe entre ces inventions une relation technique portant sur un ou plusieurs éléments techniques particuliers identiques ou correspondants. L'expression éléments techniques particuliers s'entend des éléments techniques qui déterminent une contribution de chacune des inventions revendiquées, considérée comme un tout, par rapport à l'état de la technique.

- 3.3. La relation technique reliant l'objet des **revendications 1-28** est relative au fait qu'elles concernent toutes des matériaux nucléiques dérivés d'un extrait de plasma de patients atteints de la sclérose en plaques (SEP) et dérivés d'un rétrovirus MSR/V (Multiple Sclerosis (MS) associated Retrovirus (RV)).

Cependant, cette relation ne peut pas être acceptée comme étant un "élément technique particulier", tel que défini ci-dessus, dès lors qu'elle ne détermine pas une contribution de chacune des différentes inventions revendiquées, considérée comme un tout, par rapport à l'état de la technique.

De fait, des séquences dérivées de MSR/V et se rapportant à la sclérose en plaques sont connues dans l'art antérieur.

Les documents **D1** (FR-A-2 737 500) et **D2** (WO-A-95 21256) décrivent du matériel viral, à l'état isolé ou purifié, et des fragments nucléotidiques, associés à la sclérose en plaques. Cela est reconnu par la Demanderesse, comme indiqué à la page 5 de la description de la demande. En outre, ces documents décrivent aussi l'utilisation de ces séquences à des fins de diagnostic, prophylactiques et thérapeutiques.

En conséquence, le concept général, i.e. des matériaux nucléiques dérivés d'un extrait de plasma de patients atteints de la sclérose en plaques (SEP) et dérivés d'un rétrovirus MSR/V n'est pas nouveau au vu de l'art antérieur (**Article 33 (2) PCT**).

Ainsi, compte tenu de l'art antérieur, la demande de brevet présente un défaut d'unité, dès lors qu'il est considéré que les diverses inventions ou pluralités d'inventions suivantes ne sont pas reliées entre elles de telle sorte qu'elles ne forment qu'un seul concept inventif général.

Invention 1: Revendications 5 (en entier) et 1-3, 13-19 et 22-28 (partiellement) concernant un matériel nucléique comprenant une séquence nucléotidique *SEQ ID NO: 112*, ou codant pour un polypeptide avec une séquence peptidique *SEQ ID NO: 113*, les peptides correspondants, des compositions de diagnostic, prophylactiques, ou thérapeutiques, comprenant un fragment des dites séquences et leur utilisation dans un procédé pour détecter un rétrovirus associé à la sclérose en plaques.

Invention 2: Revendications 7-10 et 12 (en entier) et 1, 2, 11 et 13-28 (partiellement) concernant un matériel nucléaire comprenant une séquence nucléotidique SEQ ID NO: 114, SEQ ID NO: 117, ou SEQ ID NO: 120 ou codant pour un polypeptide avec une séquence peptidique SEQ ID NO: 115, SEQ ID NO: 118, ou SEQ ID NO: 121, les peptides correspondants, des compositions de diagnostic, prophylactiques, ou thérapeutiques, comprenant un fragment des dites séquences et leur utilisation dans un procédé pour détecter un rétrovirus associé à la sclérose en plaques.

Invention 3: Revendications 1, 3, 13-15, 18, 19, 21-25, 27 et 28 (partiellement) concernant un matériel nucléaire comprenant une séquence nucléotidique SEQ ID NO: 124, les peptides correspondants, des compositions de diagnostic, prophylactiques, ou thérapeutiques, comprenant un fragment des dites séquences et leur utilisation dans un procédé pour détecter un rétrovirus associé à la sclérose en plaques.

Invention 4: Revendications 4 et 6 (en entier) et 1, 2, 13-19 et 21-28 (partiellement) concernant un matériel nucléaire comprenant une séquence nucléotidique SEQ ID NO: 130 ou codant pour un polypeptide avec une séquence peptidique SEQ ID NO: 135 ou SEQ ID NO: 137, les peptides correspondants, des compositions de diagnostic, prophylactiques, ou thérapeutiques, comprenant un fragment des dites séquences et leur utilisation dans un procédé pour détecter un rétrovirus associé à la sclérose en plaques.

Invention 5: Revendications 1, 11, 13-15, 18-20, 22-25, 27 et 28 (partiellement) concernant un matériel nucléaire comprenant une séquence nucléotidique SEQ ID NO: 141 ou SEQ ID NO: 142, les peptides correspondants, des compositions de diagnostic, prophylactiques, ou thérapeutiques, comprenant un fragment des dites séquences et leur utilisation dans un procédé pour détecter un rétrovirus associé à la sclérose en plaques.

- 3.4. En réponse à l'invitation à limiter les revendications ou à payer des taxes additionnelles, datée du 18.06.99, la Demanderesse a choisi de ne pas payer de taxes additionnelles et a donc limité la demande à l'**Invention 2: Revendications 7-10 et 12 (en entier) et 1, 2, 11 et 13-28 (partiellement)**.

En conséquence, le présent Rapport d'Examen Préliminaire International est donc limité à l'objet des **revendications 7-10 et 12 (en entier)** et **1, 2, 11 et 13-28 (partiellement)**.

SECTION V

4. Il est fait référence aux documents suivants:

D1: FR-A-2 737 500

D2: WO-A-95 21256

5. Activité Inventive (Article 33(3) PCT)

La présente demande ne satisfait pas aux critères énoncés à l'**Article 33 (3) PCT**, l'objet de l'**Invention 2: Revendications 7-10 et 12 (en entier)** et **1, 2, 11 et 13-28 (partiellement)** n'impliquant pas d'activité inventive (**Règle 65 (1) et (2) PCT**).

L'**état de la technique le plus proche** pour juger de l'activité inventive de l'*Invention 2* est considéré comme étant représenté par le document **D1 ou D2**.

Les documents D1 et D2, chacun pris séparément, décrivent du matériel viral, à l'état isolé ou purifié, et des fragments nucléotidiques, notamment des séquences du gène *env* du génome rétroviral MSRV-1, associés à la sclérose en plaques. Ces documents décrivent aussi des sondes nucléiques de détection qui s'hybrident auxdits lesdits fragments, des amorces comprenant tout ou partie desdits fragments et des peptides codés par lesdits fragments. En outre, ces documents décrivent l'utilisation de ces séquences à des fins de diagnostic, prophylactiques et thérapeutiques.

(Voir **D1**: Abrégé, page 2, ligne 31-page 8, ligne 5, exemples 7-9 et revendications et **D2**: Abrégé, page 3, ligne 5-page 17, ligne 14 et revendications).

La différence entre l'objet revendiqué et celui des documents D1 ou D2 réside dans les séquences spécifiques dérivées d'un rétrovirus MSRV (Multiple Sclerosis (MS) associated Retrovirus (RV)).

Au vu du document D1 ou D2, le **problème technique** à résoudre par l'*Invention 2* est celui de fournir des séquences nucléotidiques et peptidiques supplémentaires, dérivées du MSRV.

Il est de noter que le problème à résoudre n'est pas celui de déduire la séquence complète et exacte du génome réplcatif du rétrovirus.

La **solution** apportée par le Demanderesse réside dans un matériel nucléaire comprenant une séquence nucléotidique SEQ ID NO:114, SEQ ID NO:117 ou SEQ ID NO:120 et codant, respectivement, pour un polypeptide présentant les séquences peptidiques SEQ ID NO:115, SEQ ID NO:118 ou SEQ ID NO:121.

En outre, les documents **D1 et D2** indiquent que le MSRV, un virus qui contient de l'ARN, a une variabilité consécutive notamment à des taux relativement élevés de mutation spontanée. Cela implique donc que le MSRV comprend de multiples souches ou sous-familles. (Voir **D1**: page 8, lignes 16-25 et **D2**: page 15, lignes 23-32).

L'homme du métier n'aurait donc besoin d'aucune activité inventive pour utiliser les séquences décrites dans D1 et D2, qui sont déjà associées aux génomes correspondants, pour produire des amorces et/ou des sondes nucléiques pour la détection d'un agent pathogène et/ou infectant associé à la sclérose en plaques. Cela permettrait donc d'isoler des séquences supplémentaires dérivées du MSRV, qui permettraient de caractériser d'autres régions du génome de MSRV, ou des variants de MSRV.

Les séquences de l'*Invention 2* servent uniquement à une caractérisation additionnelle du système rétroviral impliqué dans la SEP (sclérose en plaques). De ce fait, les séquences de la présente demande sont des solutions alternatives pour un problème qui avait déjà été résolu de manière similaire dans l'art antérieur (D1 et D2).

Ces séquences sont considérées comme relevant d'une sélection arbitraire entre plusieurs séquences dérivées du MSRV disponibles, et qui n'implique donc à priori aucune activité inventive.

Une telle sélection relève d'une activité inventive uniquement au cas où elle produit un effet inattendu/surprenant non envisagé par l'homme du métier. Un tel effet n'a pas été démontré pour les séquences de l'*Invention 2*.

L'objet des **revendications 7-10 et 12** (en entier) et **1, 2, 11 et 13-26** (partiellement) n'implique donc pas d'activité inventive.

Dès lors que les documents D1 et D2 décrivent l'utilisation de séquences dérivées du MSRV associé à la sclérose en plaques à des fins de diagnostic, prophylactiques et thérapeutiques, ainsi que pour la détection de rétrovirus associé à la sclérose en plaques, l'objet des revendications 27 et 28 découle de manière évidente pour l'homme du métier.

En conclusion, au regard des documents cités, la matière de l'*Invention 2*. **Revendications 7-10 et 12** (en entier) et **1, 2, 11 et 13-28** (partiellement), n'implique pas d'activité inventive.

6. Application Industrielle (**Article 33(4) PCT**)

L'objet de l'*Invention 2*. **Revendications 7-10 et 12** (en entier) et **1, 2, 11 et 13-28** (partiellement) est susceptible d'application industrielle (**Article 33 (4) PCT**).

SECTION VI

7. Référence au document de brevet suivant a été faite selon la **Règle 64.3 PCT** et celui-ci est dès lors cité selon la **Règle 70.10 PCT**:

WO-A-98 23755, publié le 04.06.98, déposé le 26.11.97 et ayant la date de priorité du 26.11.96.

SECTION VIII

8.1. La présente demande ne remplit pas les conditions énoncées à l'**Article 6 PCT**, les revendications suivantes n'étant pas claires.

1. Les expressions "séquences équivalentes" (**revendications 1, 7, 14, 15, 20 et 24**) et "séquences identiques partiellement" (**revendication 24**) rendent ces revendications pas claires. Ces expressions sont vagues et imprécises dans la mesure où elles n'indiquent pas le degré d'équivalence ou d'identité, ni ne donnent aucune indication concernant les caractéristiques structurelles ou fonctionnelles de ces séquences équivalentes et/ou identiques partiellement. Ces expressions permettent donc toute interprétation individuelle.
2. L'objet de la **revendication 18** manque de clarté dû aux termes "hybrider" et "sonde". Ces termes sont vagues et imprécis, dès lors qu'ils n'indiquent pas respectivement les conditions d'hybridation et la taille de la sonde. Ce fait rend ces termes complètement ouvert à toute interprétation individuelle.
3. Le terme "partie de" rend la **revendication 20** pas claire.
Ce terme est vague et imprécis, dans la mesure où il n'indique pas la taille de cette partie, la région de la séquence nucléotidique à laquelle cette partie correspond, ni la fonction de cette partie.

Bien que la signification des termes et expressions ci-dessus mentionnés soit indiquée dans la description (pages 11-15), la signification d'une revendication doit se dégager des termes mêmes de la revendication (voir **Directives pour PCT, CIII 4.2**).

8.2. En ce qui concerne la description, la chose suivante devrait aussi être considérée:

- page 21, ligne 18: il semble que l'expression "figure 27" n'est pas correcte
- page 28, ligne 12: le numéro des SEQ ID des amorces manque.

REVENDICATIONS

- 5 1. Matériel nucléique, à l'état isolé ou purifié, comprenant une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 et
10 SEQ ID NO: 142 ; (ii) les séquences complémentaires des séquences (i) ; et (iii) les séquences équivalentes aux séquences (i) ou (ii), en particulier les séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins 70 %
15 d'homologie avec respectivement les séquences (i) ou (ii), ledit matériel nucléique étant différent de la séquence SEQ ID NO: HSAC64 du clone humain RG083M05 accessible sur la base de donnée EMBLHTG.
- 20 2. Matériel nucléique, à l'état isolé ou purifié, codant pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec une séquence peptidique choisie dans le groupe qui consiste en SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137, ledit
25 matériel nucléique étant différent de la séquence SEQ ID NO: HSAC64 du clone humain RG083M05 accessible sur la base de donnée EMBLHTG.
- 30 3. Matériel nucléique rétroviral, dont le gène pol comprend une séquence nucléotidique identique ou équivalente à une séquence choisie dans le groupe qui consiste en SEQ ID NO: 112, SEQ ID NO: 124 et leurs séquences complémentaires.
- 35 4. Matériel nucléique rétroviral, dont l'extrémité 5' du gène pol commence au nucléotide 1419 de SEQ ID NO: 130.

5. Matériel nucléaire rétroviral, dont le gène pol code pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec la séquence peptidique SEQ ID NO: 113.

6. Matériel nucléaire rétroviral, dont l'extrémité 3' du gène gag finit au nucléotide 1418 de SEQ ID NO: 130.

7. Matériel nucléaire rétroviral, dont le gène env comprend une séquence nucléotidique identique ou équivalente à une séquence choisie dans le groupe qui consiste en SEQ ID NO: 117, et ses séquences complémentaires.

8. Matériel nucléaire rétroviral, dont le gène env comprend une séquence nucléotidique qui commence au nucléotide 1 de SEQ ID NO: 117 et finit au nucléotide au nucléotide 233 de SEQ ID NO: 114.

9. Matériel nucléaire rétroviral, dont le gène env code pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec la séquence SEQ ID NO: 118.

10. Matériel nucléaire rétroviral dont la région U3R du LTR 3' comprend une séquence nucléotidique qui se termine au nucléotide 617 de SEQ ID NO: 114.

11. Matériel nucléaire rétroviral dont la région RU5 du LTR 5' comprend une séquence nucléotidique qui commence au nucléotide 755 de SEQ ID NO: 120 et finit au nucléotide 337 de SEQ ID NO: 141 ou SEQ ID NO: 142.

12. Matériel nucléaire rétroviral comprenant une séquence qui commence au nucléotide 755 de SEQ ID NO: 120 et qui se termine au nucléotide 617 de SEQ ID NO: 114.

13. Matériel nucléaire rétroviral selon l'une quelconque des revendications précédentes, caractérisé en ce qu'il est associé à au moins une maladie auto-immune telle que la sclérose en plaques ou la polyarthrite rhumatoïde.

14. Fragment nucléotidique comprenant une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, 5 SEQ ID NO: 130, SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii) les séquences complémentaires des séquences (i) ; et (iii) les séquences équivalentes aux séquences (i) ou (ii), en particulier les séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et 10 préférentiellement au moins 70 % d'homologie avec respectivement les séquences (i) ou (ii), ledit fragment nucléotidique étant différent de la séquence SEQ ID NO: HSAC64 du clone humain RG083M05 accessible sur la base de donnée EMBLHTG.

15 15. Fragment nucléotidique selon la revendication 14, consistant en une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, 20 SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii) les séquences complémentaires des séquences (i) ; et (iii) les séquences équivalentes aux séquences (i) ou (ii), en particulier les séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins 25 70 % d'homologie avec respectivement les séquences (i) ou (ii).

16. Fragment nucléotidique comprenant une séquence nucléotidique codant pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 30 50 %, et de préférence au moins 70 % d'homologie, avec une séquence peptidique choisie dans le groupe qui consiste en SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137, ledit fragment nucléotidique étant différent de la séquence 35 SEQ ID NO: HSAC64 du clone humain RG083M05 accessible sur la base de donnée EMBLHTG.

17. Fragment nucléotidique selon la revendication 16, consistant en une séquence nucléotidique codant pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec une séquence peptidique choisie dans le groupe qui consiste en SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137.

18. Sonde nucléique pour la détection d'un rétrovirus associé à la sclérose en plaques et/ou la polyarthrite rhumatoïde, caractérisée en ce qu'elle est susceptible de s'hybrider spécifiquement sur tout fragment selon l'une quelconque des revendications 14 à 17, appartenant au génome dudit rétrovirus.

19. Sonde selon la revendication 18, caractérisée en ce qu'elle possède de 10 à 100 nucléotides, de préférence de 10 à 30 nucléotides.

20. Amorce pour l'amplification par polymérisation d'un ARN ou d'un ADN d'un rétrovirus associé à la sclérose en plaques et/ou la polyarthrite rhumatoïde, caractérisée en ce qu'elle comprend une séquence nucléotidique identique ou équivalente à au moins une partie de la séquence nucléotidique d'un fragment selon l'une quelconque des revendications 14 à 17, notamment une séquence nucléotidique présentant pour toute suite de 10 monomères contigus, au moins 50 %, de préférence au moins 70 % d'homologie avec au moins ladite partie dudit fragment.

21. Amorce selon la revendication 20, caractérisée en ce que sa séquence nucléotidique est choisie parmi SEQ ID NO: 116, SEQ ID NO: 119, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 132, et SEQ ID NO: 133.

22. ARN comprenant un fragment génomique du matériel nucléique selon l'une quelconque des

revendications 1 à 7 ou un fragment selon l'une quelconque des revendications 14 à 17.

23. ADN comprenant un fragment génomique du matériel nucléaire selon l'une quelconque des revendications 1 à 7 ou un fragment selon l'une quelconque des revendications 14 à 17.

24. Vecteur de réplication et/ou d'expression, comprenant un fragment génomique du matériel nucléaire selon l'une quelconque des revendications 1 à 7 ou un fragment selon l'une quelconque des revendications 14 à 17.

25. Peptide codé par tout cadre de lecture ouvert appartenant à un fragment nucléotidique selon l'une quelconque des revendications 14 à 17, notamment polypeptide, par exemple oligopeptide formant ou comprenant un déterminant antigénique reconnu par des sera de patients infectés par le virus MSRV-1, et/ou chez lesquels le virus MSRV-1 a été réactivé.

26. Peptide selon la revendication 25 comprenant une séquence identique, partiellement ou totalement, ou équivalente à une séquence choisie parmi SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137.

27. Composition diagnostique, prophylactique, ou thérapeutique, notamment pour inhiber l'expression d'au moins un rétrovirus associé à la sclérose en plaques et/ou à la polyarthrite rhumatoïde, comprenant un fragment nucléotidique selon l'une quelconque des revendications 14 à 17.

28. Procédé pour détecter un rétrovirus associé à la sclérose en plaques et/ou à la polyarthrite rhumatoïde, dans un échantillon biologique, caractérisé en ce que l'on met en contact un ARN et/ou un ADN présumé appartenir ou provenant dudit rétrovirus, ou leur ARN et/ou ADN complémentaire, avec une composition comprenant un

fragment nucléotidique selon l'une quelconque des
revendications 14 à 17.

RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS,
IN PARTICULAR ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR
RHEUMATOID ARTHRITIS, FOR DIAGNOSTIC, PROPHYLACTIC AND
THERAPEUTIC USES

5

Multiple sclerosis (MS) is a demyelinating disease of the central nervous system (CNS) of which the complete cause still remains unknown.

Numerous studies have supported the hypothesis
10 for a viral etiology of the disease, but none of the known viruses tested has proved to be the causative agent tested for: a review of the viruses tested for in MS for many years has been carried out by E. Norrby and R.T. Johnson.

15 Recently, a retrovirus, different from the known human retroviruses, was isolated from patients suffering from MS. The authors were able to show that this retrovirus could be transmitted in vitro, that patients suffering from MS produced antibodies capable
20 of recognizing proteins associated with the infection of the leptomeningeal cells by this retrovirus, and that the expression of the latter could be greatly stimulated by the immediate-early genes of some herpesviruses.

25 All these results argue in favor of the role in MS of at least one unknown retrovirus or of a virus having a reverse transcriptase (RT) activity which is detectable by the method published by H. Perron and termed "LM7-type RT" activity.

30 The studies by the applicant have made it possible to obtain two continuous cell lines infected with natural isolates obtained from two different patients suffering from MS, by a culture method as described in the document WO-A-93 20188, whose content
35 is incorporated by reference into the present description. These two lines derived from cells of human choroid plexus, called LM7PC and PLI-2, were deposited at the E.C.A.C.C. on 22 July 1992 and 8 January 1993, respectively, under numbers 92 072201

and 93 010817, in accordance with the provisions of the Treaty of Budapest. Moreover, the viral isolates possessing an LM7-type RT activity have also been deposited at the E.C.A.C.C. under the overall name of "strains". The "strain" or isolate harbored by the PLI-2 line, called POL-2, was deposited at the E.C.A.C.C. on 22 July 1992 under No. V92072202. The "strain" or isolate harbored by the LM7PC line, called MS7PG, was deposited at the E.C.A.C.C. on 8 January 1993 under No. V93010816.

Using the abovementioned cultures and isolates, characterized by biological and morphological criteria, efforts were then made to characterize the genetic material associated with the viral particles produced in these cultures.

The proportions of genome already characterized were used to develop molecular detection tests for the viral genome and immunoserological tests, using the amino acid sequences encoded by the nucleotide sequences of the viral genome, in order to detect the immune response directed against epitopes associated with the viral infection and/or expression.

These tools have already made it possible to confirm an association between MS and the expression of the sequences identified in the patents cited further on. However, the viral system discovered by the applicant is related to a complex retroviral system. Indeed, the sequences which are found to be encapsidated in the extracellular viral particles produced by the different cultures of cells of patients suffering from MS show clearly that there is co-encapsulation of retroviral genomes which are related but different from the "wild-type" retroviral genome which produces the infectious viral particles. This phenomenon was observed between replicative retroviruses and endogenous retroviruses belonging to the same family, or even heterologous retroviruses. The concept of endogenous retrovirus is very important in the context of our discovery because, in the case of

MSRV-1, it has been observed that endogenous retroviral sequences comprising sequences homologous to the MSRV-1 genome exist in normal human DNA. The existence of endogenous retroviral elements (ERV) related to MSRV-1 through all or part of their genome explains the fact that the expression of the MSRV-1 retrovirus in human cells can interact with related endogenous sequences. These interactions are found in the case of pathogenic and/or infectious endogenous retroviruses (for example some ecotropic strains of the Murine Leukemia virus), in the case of exogenous retroviruses whose nucleotide sequence may be found partially or completely in the form of ERVs, in the genome of the host animal (e.g. mouse mammary tumor exogenous virus transmitted via milk). These interactions consist mainly of (i) a transactivation or co-activation of ERVs by the replicative retrovirus, (ii) an "illegitimate" encapsidation of related RNAs of ERVs, or of ERVs - or even of cellular RNAs - simply possessing compatible encapsidation sequences, into the retroviral particles produced by the expression of the replicative strain, which are sometimes transmissible and sometimes with an inherent pathogenicity, and (iii) relatively high recombinations between the co-encapsidated genomes, in particular in the reverse transcription phases, which lead to the formation of hybrid genomes, which are sometimes transmissible and sometimes with an inherent pathogenicity.

Thus, (i) various MSRV-1-related sequences have been found in purified viral particles; (ii) molecular analysis of the various regions of the MSRV-1 retroviral genome should be carried out by systematically analyzing the co-encapsidated, interfering and/or recombinant sequences which are generated by the infection and/or expression of MSRV-1; furthermore, some clones may have portions of defective sequences produced by the retroviral replication and the template and/or transcription errors caused by reverse transcriptase; (iii) the families of sequences

related to the same retroviral genomic region are the supports for an overall diagnostic detection which may be optimized by the identification of invariable regions among the clones expressed and by the
5 identification of reading frames responsible for the production of antigenic and/or pathogenic polypeptides which may only be produced by a portion, or even only one, of the clones expressed and under these conditions, the systematic analysis of the clones
10 expressed in one region of a given gene makes it possible to evaluate the frequency of variation and/or recombination of the MSRV-1 genome in this region and to define the optimum sequences for the applications, in particular the diagnostic applications; (iv) the
15 pathology caused by a retrovirus such as MRSV-1 may be a direct effect of its expression and of the proteins or peptides produced as a result, but also an effect of the activation, encapsidation, recombination of related or heterologous genomes and proteins or peptides
20 produced as a result; thus, these genomes associated with the expression and/or infection by MSRV-1 are an integral part of the potential pathogenicity of this virus and therefore constitute diagnostic detection supports and particular therapeutic targets. Likewise,
25 any agent which is associated with, or which is a cofactor for these interactions responsible for the pathogenicity in question, such as MSRV-2 or the gliotoxic factor described in the patent application published under the No. FR-2,716,198, can participate
30 in the development of an overall and very effective strategy for therapeutic diagnosis, prognosis, monitoring and/or integrated therapy for MS in particular, but also for any other disease associated with the same agents.

35 In this context, a parallel discovery has been made in another autoimmune disease, rheumatoid arthritis (RA), which has been described in the French⁴ patent application filed under the No. 95 029604. This discovery shows that, by applying methodological

approaches similar to those which were used in the studies by the applicant on MS, it has been possible to identify a retrovirus expressed in RA which shares the sequences described for MSRV-1 in MS and also the co-existence of an MSRV-2-associated sequence which is also described in MS. As regards MSRV-1, the sequences commonly detected in MS and RA relate to the *pol* and *gag* genes. On the basis of current knowledge, it is possible to combine the *gag* and *pol* sequences described with the MSRV-1 strains expressed in these two diseases.

The present patent application has as its object various results, supplementary in relation to those already protected by the French patent applications:

- No. 92/04322 of 03.04.1992, published under No. 2,689,519;
- No. 92/13447 of 03.11.1992, published under No. 2,689,521;
- 20 - No. 92/13443 of 03.11.1992, published under No. 2,689,520;
- No. 94/01529 of 04.02.1994, published under No. 2,715,936;
- No. 94/01531 of 04.02.1994, published under No. 2,715,939;
- 25 - No. 94/01530 of 04.02.1994, published under No. ~~2,715,936~~;
- No. 94/01532 of 04.02.1994, published under No. ~~2,715,937~~;
- 30 - No. 94/14322 of 24.11.1994, published under No. ~~2,727,428~~;
- No. 94/15810 of 23.12.1994, published under No. 2,728,585;

and

- 35 - Patent Application WO ~~97/062601~~

The present invention relates, first of all, to a nucleic material, which may consist of a retroviral material, in isolated or purified state, which may be understood or characterized in various ways:

- it comprises a nucleotide sequence chosen from the group which consists of (i) the sequences
SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117,
SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130,
5 SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences complementary to sequences (i); and (iii) the sequences equivalent to sequences (i) or (ii), in particular the sequences having, for every series of 100 contiguous monomers, at least 50%, and preferentially at least 70%
10 homology with sequences (i) or (ii) respectively;
- it encodes a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with a peptide sequence chosen from the group which consists
15 of SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137;
- its pol gene comprises a nucleotide sequence identical or equivalent to a sequence chosen from the group which consists of SEQ ID NO: 112, SEQ ID NO: 124
20 and their complementary sequences;
- the 5' end of its pol gene starts at nucleotide 1419 of SEQ ID NO: 130;
- its pol gene encodes a polypeptide having, for every contiguous series of at least 30 amino acids,
25 at least 50%, and preferably at least 70% homology with the peptide sequence SEQ ID NO: 113;
- the 3' end of its gag gene ends at nucleotide 1418 of SEQ ID NO: 130;
- its env gene comprises a nucleotide sequence
30 identical or equivalent to a sequence chosen from the group which consists of SEQ ID NO: 117, and its complementary sequences;
- its env gene comprises a nucleotide sequence which starts at nucleotide 1 of SEQ ID NO: 117 and ends
35 at nucleotide at nucleotide [sic] 233 of SEQ ID NO: 114;
- its env gene encodes a polypeptide having, for every contiguous series of at least 30 amino acids,

at least 50%, and preferably at least 70% homology with the sequence SEQ ID NO: 118;

- the U3R region of its 3' LTR comprises a nucleotide sequence which ends at nucleotide 617 of SEQ ID NO: 114;

- the RU5 region of its 5' LTR comprises a nucleotide sequence which starts at nucleotide 755 of SEQ ID NO: 120 and ends at nucleotide 337 of SEQ ID NO: 141 or SEQ ID NO: 142;

- a retroviral nucleic material comprising a sequence which starts at nucleotide 755 of SEQ ID NO: 120 and which ends at nucleotide 617 of SEQ ID NO: 114;

- the retroviral nucleic material as defined above is in particular associated with at least one autoimmune disease such as multiple sclerosis or rheumatoid arthritis.

The invention also relates to a nucleotide fragment which corresponds to at least one of the following definitions:

- it comprises or consists of a nucleotide sequence chosen from the group which consists of (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences complementary to sequences (i); and (iii) the sequences equivalent to sequences (i) or (ii), in particular the sequences having, for every series of 100 contiguous monomers, at least 50%, and preferentially at least 70% homology with sequences (i) or (ii) respectively;

- it comprises or consists of a nucleotide sequence encoding a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with a peptide sequence chosen from the group which consists of SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137.

Other subjects of the present invention are the following:

- a nucleic probe for the detection of a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, capable of hybridizing specifically with any fragment defined above and
5 belonging to the genome of said retrovirus; it advantageously possesses from 10 to 100 nucleotides, preferably from 10 to 30 nucleotides;

- a primer for the amplification, by polymerization, of an RNA or of a DNA of a retrovirus
10 associated with multiple sclerosis and/or rheumatoid arthritis, which comprises a nucleotide sequence identical or equivalent to at least a portion of the nucleotide sequence of a fragment defined above, in particular a nucleotide sequence having, for every
15 series of 10 contiguous monomers, at least 50%, preferably at least 70% homology with at least said portion of said fragment; preferably the nucleotide sequence of a primer of the invention is chosen from
SEQ ID NO: 116, SEQ ID NO: 119, SEQ ID NO: 122,
20 SEQ ID NO: 123, SEQ ID NO: 126, SEQ ID NO: 127,
SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 132, and
SEQ ID NO: 133;

- an RNA or a DNA, and in particular a replication and/or expression vector, comprising a
25 genomic fragment of the nucleic material or a fragment defined above;

- a peptide encoded by any open reading frame belonging to a nucleotide fragment defined above, in particular a polypeptide, for example oligopeptide
30 forming or comprising an antigenic determinant recognized by sera of patients infected with the MSRV-1 virus, and/or in whom the MSRV-1 virus has been reactivated; a preferential peptide comprises a sequence identical, partially or completely, or
35 equivalent to a sequence chosen from SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137;

- a diagnostic, prophylactic or therapeutic composition, in particular for inhibiting the

expression of at least one retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, comprising a nucleotide fragment defined above;

- a method for detecting a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, in a biological sample, comprising the steps consisting of bringing an RNA and/or a DNA assumed to belong to or obtained from said retrovirus, or their complementary RNA and/or DNA, into contact with a composition comprising a nucleotide fragment defined above.

Before detailing the invention, various terms used in the description and the claims are now defined:

- strain or isolate is understood to mean any infectious and/or pathogenic biological fraction containing, for example, viruses and/or bacteria and/or parasites, generating a pathogenic and/or antigenic power, harbored by a culture or a live host; by way of example, a viral strain according to the preceding definition may contain a co-infectious agent, for example a pathogenic protist,

- the term "MSRV" used in the present description designates any pathogenic and/or infectious agent, as associated with MS, in particular a viral species, the attenuated strains of said viral species, or the interfering defective particles or particles containing co-encapsidated genomes or alternatively genomes recombined with a portion of the MSRV-1 genome, which are derived from this species. It is known that viruses and particularly viruses containing RNA exhibit variability, following in particular relatively high rates of spontaneous mutation, which will be taken into account below to define the concept of equivalence,

- human virus is understood to mean a virus capable of infecting or of being harbored by human beings,

- given all the natural or induced variations and/or recombination which may be encountered in practice in the present invention, the objects thereof,

defined above and in the claims, have been expressed by comprising the equivalents or derivatives of the various biological materials defined below, in particular homologous nucleotide or peptide sequences,

5 - the variant of a virus or of a pathogenic and/or infectious agent according to the invention comprises at least one antigen recognized by at least one antibody directed against at least one corresponding antigen of said virus and/or of said
10 pathogenic and/or infectious agent, and/or a genome in which any portion is detected by at least one hybridization probe, and/or at least one nucleotide amplification primer specific for said virus and/or pathogenic and/or infectious agent, under defined
15 hybridization conditions well known to persons skilled in the art,

 - according to the invention, a nucleotide fragment or an oligonucleotide or a polynucleotide is a stretch of monomers, or a biopolymer, characterized by
20 the informational sequence of the natural nucleic acids, which is capable of hybridizing to any other nucleotide fragment under predefined conditions, it being possible for the stretch to contain monomers of different chemical structures and to be obtained from a
25 natural nucleic acid molecule and/or by genetic recombination and/or by chemical synthesis; a nucleotide fragment may be identical to a genomic fragment of the MSRV-1 virus considered by the present invention, in particular a gene of the latter, for
30 example pol or env in the case of said virus;

 - thus, a monomer may be a natural nucleic acid nucleotide in which the constituent components are a sugar, a phosphate group and a nitrogen base; in RNA, the sugar is ribose; in DNA, the sugar is 2-deoxy-
35 ribose; depending on whether DNA or RNA is involved, the nitrogen base is chosen from adenine, guanine, uracil, cytosine, thymine; or the nucleotide may be modified in at least one of the three constituent components; by way of example, the modification may

occur at the level of the bases, generating modified bases such as inosine, 5-methyl-deoxycytidine, deoxyuridine, 5-dimethylamineodeoxyuridine [sic], 2,6-diamineopurine [sic], 5-bromodeoxyuridine and any
5 other modified base promoting hybridization; at the level of the sugar, the modification may consist in the replacement of at least one deoxyribose with a polyamide, and at the level of the phosphate group, the modification may consist in its replacement with
10 esters, in particular chosen from the esters of diphosphate, of alkyl and arylphosphonate and of phosphorothioate,

- "informational sequence" is understood to mean any ordered series of monomers, whose chemical
15 nature and in which the order in a reference direction, constitute or otherwise a functional information of the same quality as that for the natural nucleic acids,

- hybridization is understood to mean the process during which, under appropriate operating
20 conditions, two nucleotide fragments, having sufficiently complementary sequences, become annealed to form a complex, in particular a double or triple, structure, preferably in helical form,

- a probe comprises a nucleotide fragment
25 synthesized by the chemical route or obtained by digestion or enzymatic cleavage of a longer nucleotide fragment, comprising at least six monomers, advantageously from 10 to 100 monomers, preferably 10 to 30 monomers, and possessing a hybridization
30 specificity under defined conditions; preferably, a probe possessing less than 10 monomers is not used alone, but is used in the presence of other probes which are equally short in length or otherwise; under certain specific conditions, it may be useful to use
35 probes which are greater than 100 monomers in size; a probe may be used in particular for diagnostic purposes, and it may be, for example, capture and/or detection probes,

- the capture probe may be immobilized on a solid support by any appropriate means, that is to say directly or indirectly, for example by covalent bonding or passive adsorption,

5 - the detection probe may be labeled by means of a marker chosen in particular from radioactive isotopes, enzymes chosen in particular from peroxidase and alkaline phosphatase and those capable of hydrolyzing a chromogenic, fluorogenic or luminescent
10 substrate, chromophoric chemical compounds, chromogenic, fluorogenic or luminescent compounds, analogs of nucleotide bases, and biotin,

 - the probes used for diagnostic purposes of the invention may be used in all known hybridization
15 techniques, and in particular the so-called "DOT-BLOT" technique, "SOUTHERN BLOT" technique, "NORTHERN BLOT" technique which is a technique identical to the "SOUTHERN BLOT" technique but which uses RNA as target, the SANDWICH technique; advantageously, the SANDWICH
20 technique is used in the present invention, comprising a specific capture probe and/or a specific detection probe, it being understood that the capture probe and the detection probe must have a nucleotide sequence which is at least partially different,

25 - any probe according to the present invention may hybridize in vivo or in vitro with the RNA and/or with the DNA, in order to block the replication, in particular translation and/or transcription, phenomena and/or to degrade said DNA and/or RNA,

30 - a primer is a probe comprising at least six monomers, and advantageously from 10 to 30 monomers, possessing hybridization specificity under defined conditions, for the initiation of an enzymatic polymerization, for example in an amplification
35 technique such as PCR (Polymerase Chain Reaction), in an extension method such as sequencing, in a reverse transcription method and the like,

 - two nucleotide or peptide sequences are said to be equivalent or derived with respect to each other,

or with respect to a reference sequence, if functionally the corresponding biopolymers can play substantially the same role, without being identical, in relation to the application or use considered, or in
5 the technique in which they are involved; particularly equivalent are two sequences obtained because of the natural variability, in particular spontaneous mutation, of the species from which they were identified, or induced mutation, as well as two
10 homologous sequences, the homology being defined below,

- "variability" is understood to mean any spontaneous or induced modification of a sequence, in particular by substitution, and/or insertion, and/or deletion of nucleotides and/or of nucleotide fragments,
15 and/or extension and/or shortening of the sequence at least at one of the ends; a nonnatural variability may result from the genetic engineering techniques used, for example from the choice of the degenerate or nondegenerate synthetic primers selected to amplify a
20 nucleic acid; this variability may result in modifications of any starting sequence, considered as a reference, and which may be expressed by a degree of homology with respect to said reference sequence,

- homology characterizes the degree of identity
25 of two compared nucleotide or peptide fragments; it is measured by the percentage identity which is in particular determined by direct comparison of nucleotide or peptide sequences, with respect to reference nucleotide or peptide sequences,

30 - any nucleotide fragment is said to be equivalent to or derived from a reference fragment if it has a nucleotide sequence equivalent to the sequence of the reference fragment; according to the preceding definition, in particular equivalent to a reference
35 nucleotide fragment are:

(a) any fragment capable of hybridizing, at least partially, with the complementary to the reference fragment,

(b) any fragment whose alignment with the reference fragment leads to the identification of identical contiguous bases, in a greater number than with any other fragment obtained from another taxonomic group,

(c) any fragment resulting or capable of resulting from the natural variability of the species from which it is obtained,

(d) any fragment which may result from genetic engineering techniques applied to the reference fragment,

(e) any fragment, containing at least eight contiguous nucleotides, encoding a peptide homologous or identical to the peptide encoded by the reference fragment,

(f) any fragment different from the reference fragment through insertion, deletion, substitution of at least one monomer, extension, or shortening at least at one of its ends; for example, any fragment corresponding to the reference fragment, flanked at least at one of its ends by a nucleotide sequence not encoding a polypeptide,

- polypeptide is understood to mean in particular any peptide of at least two amino acids, in particular oligopeptide, protein, extracted, separated, or substantially isolated or synthesized, through the involvement of humans, in particular those obtained by chemical synthesis, or through expression in a recombinant organism,

- polypeptide partially encoded by a nucleotide fragment is understood to mean a polypeptide having at least three amino acids encoded by at least nine contiguous monomers included in said nucleotide fragment,

- an amino acid is said to be analogous to another amino acid when their respective physicochemical characteristics, such as polarity, hydrophobicity and/or basicity, and/or acidity, and/or

neutrality, are substantially the same; thus, a leucine is analogous to an isoleucine,

- any polypeptide is said to be equivalent to or derived from a reference polypeptide if the polypeptides compared have substantially the same properties, and in particular the same antigenic, immunological, enzymatic and/or molecular recognition properties; in particular equivalent to a reference polypeptide is:

(a) any polypeptide possessing a sequence in which at least one amino acid has been replaced by an analogous amino acid,

(b) any polypeptide having an equivalent peptide sequence, obtained by natural or induced variation of said reference polypeptide, and/or of the nucleotide fragment encoding said polypeptide,

(c) a mimotope of said reference polypeptide,

(d) any polypeptide from whose sequence one or more amino acids of the L series are replaced by an amino acid of the D series, and vice versa,

(e) any polypeptide into whose sequence a modification of the side chains of the amino acids has been introduced, such as for example an acetylation of the amine-containing functions, a carboxylation of the thiol functions, an esterification of the carboxyl functions,

(f) any polypeptide in whose sequence one or more peptide bonds have been modified, such as for example the carba, retro, inverso, retro-inverso, reduced, and methylene-oxy bonds,

(g) any polypeptide in which at least one antigen is recognized by an antibody directed against a reference polypeptide,

- the percentage identity characterizing the homology between two peptide fragments compared is according to the present invention at least 50% and preferably at least 70%.

Given that a virus possessing a reverse transcriptase enzymatic activity may be genetically

characterized both in RNA and DNA form, both the viral DNA and RNA will be mentioned in order to characterize the sequences relative to a virus possessing such a reverse transcriptase activity, termed MSRV-1 according to the present description.

The expressions of order which are used in the present description and the claims, such as "first nucleotide sequence", are not selected to express a particular order, but to define the invention more clearly.

Detection of a substance or agent is understood below to mean an identification, a quantification or a separation or isolation of said substance or of said agent.

The invention will be understood more clearly on reading the detailed description which follows which is made with reference to the appended figures in which:

Figure 1 represents the general structure of the proviral DNA and the genomic RNA of MSRV-1.

Figure 2 represents the nucleotide sequence of the clone called CL6-5' (SEQ ID NO: 112) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 3 represents the nucleotide sequence of the clone called CL6-3' (SEQ ID NO: 114) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 4 represents the nucleotide sequence of the clone called C15 (SEQ ID NO: 117) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 5 represents the nucleotide sequence of the clone called 5M6 (SEQ ID NO: 120) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 6 represents the nucleotide sequence of the clone called CL2 (SEQ ID NO: 130) and three

potential reading frames in amino acids presented under the nucleotide sequence.

Figure 7 represents three potential reading frames in amino acids expressed by pET28C-clone 2 and presented under the nucleotide sequence.

Figure 8 represents three potential reading frames in amino acids expressed by pET21C-clone 2 and presented under the nucleotide sequence.

Figure 9 represents the nucleotide sequence of the clone called LB13 (SEQ ID NO: 141) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 10 represents the nucleotide sequence of the clone called LA15 (SEQ ID NO: 142) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 11 represents the nucleotide sequence of the clone called LB16 (SEQ ID NO: 124) and three potential reading frames in amino acids presented under the nucleotide sequence.

Figure 12 represents the promoter activity expressed in cpm/4 min of the U3R sequences subcloned from LTRs of different origins into the plasmid PCAT3. PCAT3 means plasmid alone, PCAT-PH74 means plasmid plus endogenous U3R clone expressed in the placenta, PCAT-cl6 means plasmid plus U3R clone amplified in the RNA of an MS plasma, PCAT-5M6 means plasmid plus U3R region amplified in the cellular DNA, "no plasmid" means absence of plasmid in the test.

Figure 13 represents the MSRV1 env and 3' LTR sequences. The horizontal arrows indicate the start of the env, U3 and R regions. In the env region, the signal peptide and the potential immunosuppressive region are underlined, the potential glycosilation sites are boxed and the potential cleavage sites are indicated by vertical arrows. In the U3R region: the regulatory element CAAT and the TATA Box are underlined, the "cap" site and the polyadenylation signal are also indicated.

Figure 14 represents the 5' LTR (RU5) region followed by a PBS site (primer binding site) complementary to the Trp tRNA and by a gag gene encoding a protein of about 487 amino acids. The amino acids conserved in the nucleocapsid are underlined twice. The amino acids defining the region of greatest homology in the capsid are in bold and underlined once. The / symbols in the amino acid sequence indicate variations observed depending on the clones and, in the nucleotide sequence, they indicate frame jumps in some clones. The boxed regions correspond to epitopes identified by peptide analysis of the C-terminal region.

Figure 15 represents the integrase region of MSRV1, the nucleotide sequence and the amino acid sequence deduced from the integrase region corresponding to clone 87-23. In Figure 15, // means a frame jump which has been suppressed in order to restore the potential ORF. The letters in underlined bold characters represent the conserved amino acids in the retroviral integrases.

Figure 16 describes the nucleotide and peptide sequences of clone B13 (identical to clone FBd13 described in previous applications) with indication of the ORFs and stop codons represented by a dot. The underlined region in bold represents the potential immunosuppressive domain. The single underlined domain represents the start of the 3' LTR.

EXAMPLE 1: PREPARATION OF A CL6-5' REGION ENCODING THE N-TERMINAL END OF INTEGRASE AND OF A CL6-3' REGION CONTAINING THE 3' TERMINAL SEQUENCE OF THE MSRV-1 GENOME

A 3' RACE was carried out on the total RNA extracted from plasma from a patient suffering from MS. A healthy control plasma, treated under the same conditions, was used as negative control. The synthesis of cDNA was carried out with an oligo dT primer identified by SEQ ID NO: 68 (5' GAC TCG CTG CAG ATC GAT

TTT TTT TTT TTT TTT T 3') and the reverse transcriptase "ExpandTM RT" from Boehringer according to the conditions recommended by the company. A PCR was carried out with the enzyme Klentaq (Clontech) under
5 the following conditions: 94°C 5 min then 93°C 1 min, 58°C 1 min, 68°C 3 min over 40 cycles and 68°C for 8 min, with a final reaction volume of 50 µl.

Primers used for the PCR:

- 5' primer, identified by SEQ ID NO: 69
- 10 5' GCC ATC AAG CCA CCC AAG AAC TCT TAA CTT 3';
- 3' primer, identified by SEQ ID NO: 68

A second so-called "seminested" PCR was carried out with a 5' primer situated inside the region already amplified. This second PCR was carried out under the
15 same experimental conditions as those used for the first PCR, using 10 µl of the amplification product derived from the first PCR.

Primers used for the seminested PCR:

- 5' primer, identified by SEQ ID NO: 70
- 20 5' CCA ATA GCC AGA CCA TTA TAT ACA CTA ATT 3';
- 3' primer, identified by SEQ ID NO: 68

The primers SEQ ID NO: 69 and SEQ ID NO: 70 are specific for the pol region of MRSV-1.

An amplification product of 1.9 Kb was obtained
25 for the plasma of the MS patient. The corresponding fragment was not observed for the healthy control plasma. This amplification product was cloned in the following manner:

The amplified DNA was inserted into a plasmid with the
30 aid of the TA Cloning kit[®]. The 2 µl of DNA solution were mixed with 5 µl of sterile distilled water, 1 µl of a 10 times concentrated ligation buffer "10X LIGATION BUFFER", 2 µl of "PCRTM VECTOR" (25 ng/ml) and 1 µl of "T4 DNA LIGASE". This mixture was incubated
35 overnight at 12°C. The next steps were carried out in accordance with the instructions for the TA Cloning kit[®] (Invitrogen). At the end of the procedure, the white colonies of recombinant bacteria (white) were subcultured so as to be cultured and allow the

extraction of the plasmids incorporated according to the so-called "miniprep" procedure. The plasmid preparation of each recombinant colony was cut with an appropriate restriction enzyme and analyzed on agarose
5 gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the Sp6 promoter present on the cloning plasmid of the TA
10 cloning kit[®]. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM[™] Ready Reaction AmpliTaq[®] FS, DyeDeoxy[™] Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing
15 was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The clone obtained contains a CL6-5' region encoding the N-terminal end of integrase and a CL6-3'
20 region corresponding to the 3' terminal region of MSRV-1 and making it possible to define the end of the envelope (234 bp) and the U3 and R (401 bp) regions of the MSRV1 retrovirus.

The region corresponding to the N-terminal end
25 of integrase is represented by its nucleotide sequence (SEQ ID NO: 112) in Figure 27. The three potential reading frames are presented by their amineo [sic] acid sequence under the nucleotide sequence, and the amineo [sic] acid sequence of the N-terminal end of integrase
30 is identified by SEQID NO: 113.

The C16-3' region is represented by its nucleotide sequence (SEQ ID NO: 114) in Figure 3. The three potential reading frames are presented by their amineo [sic] acid sequence under the nucleotide
35 sequence. An amineo [sic] acid sequence corresponding to the C-terminal end of the MSRV-1 env protein is identified by SEQ ID NO: 115.

In order to evaluate the promoter activity of the LTR obtained from clone 6 (cl6), a test of promoter

activity using the enzyme CAT (chloramphenicol acetyl transferase) was carried out with the corresponding U3R region. In parallel, a clone containing the same U3R region of endogenous retroviral RNA expressed in normal placenta (PH74) and a clone (5M6) obtained from DNA were tested. The result presented in Figure 12 shows a very high promoter activity of the LTR derived from MS plasma (cl6) and a significantly much lower activity with the sequences of non-MS endogenous origin.

10

EXAMPLE 2: PREPARATION OF THE C15 CLONE CONTAINING THE REGION ENCODING A PORTION OF THE MSRV-1 RETROVIRUS ENVELOPE

15 A RT-PCR was carried out on the total RNA extracted from virions concentrated by ultracentrifugation of a synoviocyte culture supernatant obtained from an MS patient. The synthesis of cDNA was carried out with an oligo dT primer and the reverse transcriptase "Expand™ RT" from Boehringer according to the conditions recommended by the company. A PCR was carried out with the Expand™ Long Template PCR System (Boehringer) under the following conditions: 94°C 5 min then 93°C 1 min, 60°C 1 min, 68°C 3 min over 40 cycles and 68°C for 8 min and with a final reaction volume of 25 50 µl.

Primers used for the PCR:

- 5' primer, identified by SEQ ID NO: 69
5' GCC ATC AAG CCA CCC AAG AAC TCT TAA CTT 3';
- 3' primer, identified by SEQ ID NO: 116
30 5' TGG GGT TCC ATT TGT AAG ACC ATC TGT AGC TT 3'

A second so-called "seminested" PCR was carried out with a 5' primer situated inside the region already amplified. This second PCR was carried out under the same experimental conditions as those used for the first PCR (except that 30 cycles were used instead of 40), using 10 µl of the amplification product derived from the first PCR.

Primers used for the seminested PCR:

- 5' primer, identified by SEQ ID NO: 70

5' CCA ATA GCC AGA CCA TTA TAT ACA CTA ATT 3';

- 3' primer, identified by SEQ ID NO: 116

The primers SEQ ID NO: 69 and SEQ ID NO: 70 are specific for the pol region of MRSV-1. The primer SEQ ID NO: 116 is specific for the sequence FBd13 (also called B13) and is located in the conserved env region among the oncoretroviruses.

An amplification product of 1932 bp was obtained and cloned in the following manner:

the amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit[®]. The various steps were carried out in accordance with the instructions for the TA Cloning kit[®] (Invitrogen). At the end of the procedure, the white colonies of recombinant bacteria (white) were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "miniprep" procedure. The plasmid preparation of each recombinant colony was cut with an appropriate restriction enzyme and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the SP6 promoter present on the cloning plasmid of the TA cloning kit[®]. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM[™] Ready Reaction AmpliTaq[®] FS, DyeDeoxy[™] Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The C15 clone obtained contains a region corresponding to the region of the MSRV-1 envelope of 1481 bp.

The env region of the C15 clone is represented by its nucleotide sequence (SEQ ID NO: 117) in Figure 5. The three potential reading frames of this clone are presented by their amino acid sequence

under the nucleotide sequence. The reading frame corresponding to an MSRV-1 structural env protein is identified by SEQ ID NO: 118.

From the defined sequences obtained from clones
5 cl6 and C15, it was possible to produce a plasmid construct encoding a complete envelope followed by the 3' LTR, as presented in Figure 13 with the corresponding reading frame.

10 **EXAMPLE 3**: PREPARATION OF A 5M6 CLONE CONTAINING THE SEQUENCES OF THE 3' TERMINAL REGION OF THE ENVELOPE, FOLLOWED BY THE MSRV-1 PROVIRAL TYPE U3, R AND U5 SEQUENCES

A monodirectional PCR was carried out on the
15 DNA extracted from immortalized B lymphocytes in culture from an MS patient. The PCR was carried out with Expand™ Long Template PCR System (Boehringer) under the following conditions: 94°C 3 min then 93°C 1 min, 60°C 1 min, 68°C 3 min over 10 cycles, then 93°C
20 1 min, 60°C 1 min with 15 sec of extension at each cycle, 68°C 3 min over 35 cycles and 68°C for 7 min and with a final reaction volume of 50 µl.

The primer used for the PCR identified by SEQ ID NO: 119 is 5' TCA AAA TCG AAG AGC TTT AGA CTT GCT
25 AAC CG 3';

The primers [sic] SEQ ID NO: 119 is specific for the env region of the C15 clone.

An amplification product of 1673 bp was obtained and cloned in the following manner:
30 the amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit®. The various steps were carried out in accordance with the instructions for the TA Cloning kit® (Invitrogen). At the end of the procedure, the white colonies of recombinant bacteria
35 (white) were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "miniprep" procedure. The plasmid preparation of each recombinant colony was cut with an appropriate restriction enzyme and analyzed on agarose

gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid of the TA cloning kit[®]. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM[™] Ready Reaction AmpliTaq[®] FS, DyeDeoxy[™] Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The 5M6 clone obtained contains a region corresponding to the 3' region of the MSRV-1 envelope of 492 bp followed by the regions U3, R and U5 (837 bp) of MSRV1.

The 5M6 clone is represented by its nucleotide sequence (SEQ ID NO: 120) in Figure 5. The three potential reading frames of this clone are presented by their amineo [sic] acid sequence under the nucleotide sequence. The reading frame corresponding to the C-terminal end of the MSRV-1 env protein is identified by SEQ ID NO: 121.

25

EXAMPLE 4: PREPARATION OF THE LB16 CLONE CONTAINING THE REGION ENCODING THE MSRV-1 RETROVIRUS INTEGRASE

An RT-PCR was carried out on the total RNA treated with DNaseI and extracted from a choroid plexus obtained from an MS patient. The synthesis of cDNA was carried out with an oligo dT primer and the reverse transcriptase "Expand[™] RT" from Boehringer according to the conditions recommended by the company. A "no RT" control was carried out in parallel on the same material. A PCR was carried out with Taq polymerase (Perkin Elmer) under the following conditions: 95°C 5 min, then 95°C 1 min, 55°C 1 min, 72°C 2 min over 35 cycles and 72°C for 8 min and with a final reaction volume of 50 µl.

Primers used for the PCR:

- 5' primer, identified by SEQ ID NO: 122

5' GGC ATT GAT AGC ACC CAT CAG 3';

- 3' primer, identified by SEQ ID NO: 123

5 5' CAT GTC ACC AGG GTG GAA TAG 3'

The primer SEQ ID NO: 122 is specific for the pol region of MSRV-1 and more precisely similar to the integrase region described above. The primer SEQ ID NO 123 was defined on sequences of the clones obtained during preliminary tests.

An amplification product of about 760 bp was obtained only in the test with RT and was cloned in the following manner:

the amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit[®]. The various steps were carried out in accordance with the instructions for the TA Cloning kit[®] (Invitrogen). At the end of the procedure, the white colonies of recombinant bacteria (white) were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "miniprep" procedure. The plasmid preparation of each recombinant colony was cut with an appropriate restriction enzyme and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid of the TA cloning kit[®]. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM[™] Ready Reaction AmpliTaq[®] FS, DyeDeoxy[™] Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The LB16 clone obtained contains the sequences corresponding to integrase. The nucleotide sequence of

this clone was identified by SEQ ID NO: 124 in Figure 11, three reading frames are determined.

EXAMPLE 5: PREPARATION OF A CLONE 2, CL2, CONTAINING IN 3' A PORTION HOMOLOGOUS TO THE POL GENE, CORRESPONDING TO THE PROTEASE GENE, AND TO THE GAG GENE (GM3) CORRESPONDING TO THE NUCLEOCAPSID, AND A NEW 5' CODING REGION, CORRESPONDING TO THE GAG GENE MORE SPECIFICALLY THE TEMPLATE AND THE CAPSID of MSRV-1.

A PCR amplification was carried out on the total RNA extracted from 100 µl of plasma from a patient suffering from MS. A water control, treated under the same conditions, was used as negative control. The synthesis of cDNA was carried out with 300 pmol of a random primer (GIBCO-BRL, France) and the reverse transcriptase "Expand RT" (BOEHRINGER MANNHEIM, France) according to the conditions recommended by the company. An amplification by PCR ("polymerase chain reaction") was carried out with the enzyme Taq polymerase (Perkin Elmer, France) using 10 µl of cDNA under the following conditions: 94°C 2 min, 55°C 1 min, 72°C 2 min then 94°C 1 min, 55°C 1 min, 72°C 2 min over 30 cycles and 72°C for 7 min with a final reaction volume of 50 µl.

Primers used for the PCR amplification:

- 5' primer, identified by SEQ ID NO: 126
5' CGG ACA TCC AAA GTG ATG GGA AAC G 3';
- 3' primer, identified by SEQ ID NO: 127
5' GGA CAG GAA AGT AAG ACT GAG AAG GC 3'

A second amplification by so-called "seminested" PCR was carried out with a 5' primer situated inside the region already amplified. This second PCR was carried out under the same experimental conditions as those used during the first PCR, using 10 µl of the amplification product derived from the first PCR.

Primers used for the amplification by seminested PCR:

- 5' primer, identified by SEQ ID NO: 128
5' CCT AGA ACG TAT TCT GGA GAA TTG GG 3';

- 3' primer, identified by SEQ ID NO: 129
5' TGG CTC TCA ATG GTC AAA CAT ACC CG 3'

The primers SEQ ID NO: [lacuna] and SEQ ID NO: [lacuna] are specific for the pol region, clone G+E+A, more specifically the E region: nucleotide position No. 423 to No. 448. The primers used in the 5' region were defined on sequences of clones obtained during preliminary tests.

An amplification product of 1511 bp was obtained from the RNA extracted from the plasma of an MS patient. The corresponding fragment was not observed for the water control. This amplification product was cloned in the following manner.

The amplified DNA was inserted into a plasmid with the aid of the TA Cloning kit™. The 2 µl of DNA solution were mixed with 5 µl of sterile distilled water, 1 µl of a 10 times concentrated ligation buffer "10X LIGATION BUFFER", 2 µl of "PCR™ VECTOR" (25 ng/ml) and 1 µl of "T4 DNA LIGASE". This mixture was incubated overnight at 14°C. The following steps were carried out in accordance with the instructions of the TA Cloning kit® (Invitrogen). The mixture was plated after transformation of the ligation into *E. coli* INVαF' bacteria. At the end of the procedure, the white colonies of recombinant bacteria were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "DNA minipreparation" procedure (17). The plasmid preparation of each recombinant colony was cut with the restriction enzyme EcoRI and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid of the TA cloning kit®. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISM™ Ready Reaction Amplitaq® FS, DyeDeoxy™ Terminator" (Applied

Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

5 The clone obtained, called CL2, contains a C-terminal region similar to the 5' terminal region of the clones G+E+A of MSRV-1, which makes it possible to define the C-terminal region of the gag gene and a new region corresponding to the N-terminal region of the
10 MSRV-1 gag gene.

 CL2 makes it possible to define a region of 1511 bp having an open reading frame in the N-terminal region of 1077 bp encoding 359 amino acids and a non-open reading frame of 454 bp corresponding to the
15 C-terminal region of the MSRV-1 gag gene.

 The nucleotide sequence of CL2 is identified by SEQ ID NO: 130. It is represented in Figure 6 with the potential reading frames in amineo [sic] acid.

 The 1077 bp fragment of CL2 encoding 359 amino
20 acids was amplified by PCR with the Pwo enzyme (5U/μl) (Boehringer Mannheim, France) using 1 μl of the DNA minipreparation of clone 2 under the following conditions: 95°C 1 min, 60°C 1 min, 72°C 2 min over 25 cycles and with a final reaction volume of 50 μl
25 with the aid of the primers:

- 5' primer (*Bam*HI), identified by SEQ ID NO: 132
5' TGC TGG AAT TCG GGA TCC TAG AAC GTA TTC 3' (30 mer),
and

- 3' primer (*Hind*III), identified by SEQ ID NO: 133
30 5 AGT TCT GCT CCG AAG CTT AGG CAG ACT TTT 3' (30 mer)
corresponding, respectively, to the nucleotide sequence of clone 2 at position -9 to 21 and 1066 to 1095.

 The fragment obtained by PCR was linearized with *Bam*HI and *Hind*III and subcloned into the
35 expression vectors pET28C and pET21C (NOVAGEN) linearized with *Bam*HI and *Hind*III. The sequencing of the DNA of the 1077 bp fragment of clone 2 in the two expression vectors was carried out according to the method recommended for the use of the sequencing kit

"PRISM™ Ready Reaction Amplitaq® FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the
5 manufacturer's instructions.

The expression of the nucleotide sequence of the 1077 bp fragment of clone 2 by the expression vectors pET28C and pET21C are identified by SEQ ID NO: 135 and SEQ ID NO: 137, respectively.

10

EXAMPLE 6: EXPRESSION OF CLONE 2 IN *ESCHERICHIA COLI*

The constructs pET28c-clone 2 (1077 bp) and pET21C-clone 2 (1077 bp) synthesize, in the bacterial strain BL21 (DE3), a protein fused at the N- and C-terminus for the vector pET28C and the C-terminus for
15 the vector pET21C with 6 Histidines, having an apparent molecular mass of about 45 kDa, identified by SDS-PAGE polyacrylamide gel electrophoresis (SDS = Sodium Dodecyl Sulfate) (Laemmli, 1970 (1)). The reactivity of
20 the protein was demonstrated towards an anti-Histidine monoclonal antibody (DIANOVA) by the Western-blot technique (Towbin et al., 1979 (2)).

The recombinant proteins pET28c-clone 2 (1077 bp) and pET21C-clone 2 (1077 bp) were visualized
25 by SDS-PAGE in the insoluble fraction after enzymatic digestion of the bacterial extracts with 50 µl of lysozyme (10 mg/ml) and ultrasound lysis.

The antigenic properties of the recombinant antigens pET28C-clone 2 (1077 bp) and pET21C-clone 2
30 (1077 bp) were tested by Western blotting () [sic] after solubilization of the bacterial pellet with 2% SDS and 50 mM β-mercaptoethanol. After incubation with sera from patients suffering from multiple sclerosis, the sera from neurological controls and the sera from
35 controls at the Blood Transfusion Center (CTS), the immunocomplexes were detected with the aid of an alkaline phosphatase-coupled goat serum anti-human IgG and anti-human IgM.

The results are presented in the table below.

TABLE

Reactivity of sera affected by multiple sclerosis and controls with the MSRV-1 recombinant protein gag clone 2 (1077 bp) = pET21C-clone 2 (1077 bp) and pET28C-clone 2 (1077 bp)^a

DISEASE	NUMBER OF INDIVIDUALS TESTED	NUMBER OF POSITIVE INDIVIDUALS
MS	15	6 2(+++), 2(++), (2(+))
NEUROLOGICAL CONTROLS	2	1(+++)
HEALTHY CONTROLS (CTS)	22	1(+/-)

(a) The strips containing 1.5 µg of recombinant antigen pET-gag clone 2 (1077 bp) exhibit reactivity against sera diluted 1/100. The Western-Blot interpretation is based on the presence or absence of a specific pET-gag clone 2 (1077 bp) band on the strips. Positive and negative controls are included in each experiment.

These results show that, under the technical conditions used, about 40% of the human sera affected by multiple sclerosis which were tested react with the recombinant proteins pET28C-clone 2 (1077 bp) and pET21C-clone 2 (1077 bp). Reactivity was observed on a neurological control and it is of interest to note that the RNAs extracted from this serum, after the reverse transcriptase step, are also amplified by PCR in the pol region. This suggests that people who have not declared MS may also harbor and express this virus. On the other hand, an apparently healthy control (CTS donor) possesses anti-gag (clone 2, 1077 bp) antibodies. This is compatible with an immunity acquired against MSRV-1 independently of a declared associated autoimmune disease.

EXAMPLE 7: PREPARATION OF AN LB13 CLONE CONTAINING IN 3' A PORTION HOMOLOGOUS TO CLONE 2 CORRESPONDING TO THE GAG GENE AND IN 5' A PORTION HOMOLOGOUS TO THE 5M6 CLONE CORRESPONDING TO THE U5 LTR REGION

5 An RT-PCR ("reverse transcriptase-polymerase chain reaction") was carried out using total RNA extracted from virions, obtained from supernatants of B lymphocyte cells of patients suffering from multiple sclerosis, concentrated by ultracentrifugations. The
10 synthesis of cDNA was carried out with a specific primer SEQ No. XXX and the reverse transcriptase "Expand™ RT" from BOEHRINGER MANNHEIM according to the conditions recommended by the company.

Primer used for the synthesis of the cDNA, identified
15 by SEQ ID NO: 138:

5' CTT GGA GGG TGC ATA ACC AGG GAA T 3'

A PCR amplification was carried out with *Tag* polymerase (Perkin Elmer, France) under the following conditions: 94°C 1 min, 55°C 1 min, 72°C 2 min over 35
20 cycles at 72°C for 7 min and with a final reaction volume of 100 µl.

Primers used for the PCR amplification:

- 5' primer, identified by SEQ ID NO: 139

5' TGT CCG CTG TGC TCC TGA TC 3'

25 - 3' primer, identified by SEQ ID NO: 138

5' CTT GGA GGG TGC ATA ACC AGG GAA T 3'

A second so-called "seminested" PCR
amplification was carried out with a 3' primer situated inside the region already amplified. This second
30 amplification was carried out under the same experimental conditions as those used during the first amplification, using 10 µl of the amplification product derived from the first PCR.

Primers used for the "seminested" PCR amplification:

35 - 5' primer, identified by SEQ ID NO: 139

5' TGT CCG CTG TGC TCC TGA TC 3'

- 3' primer, identified by SEQ ID NO: 140

5' CTA TGT CCT TTT GGA CTG TTT GGG T 3'

The primers SEQ ID NO: 138 and SEQ ID NO: 140 are specific for the gag region, clone 2 nucleotide position No. 373-397 and No. 433-456. The primers used in the 5' region were defined on sequences of the clones obtained during preliminary tests.

An amplification product of 764 bp was obtained and cloned in the following manner:

The amplified DNA was inserted into a plasmid with the aid of the TA Cloning kitTM. The 2 µl of DNA solution were mixed with 5 µl of sterile distilled water, 1 µl of a 10 times concentrated ligation buffer "10X LIGATION BUFFER", 2 µl of "pCRTM VECTOR" (25 ng/ml) and 1 µl of "T4 DNA LIGASE". This mixture was incubated overnight at 14°C. The following steps were carried out in accordance with the instructions of the TA Cloning kit[®] (Invitrogen). The mixture was plated after transformation of the ligation into *E. coli* INVαF' bacteria. At the end of the procedure, the white colonies of recombinant bacteria were subcultured so as to be cultured and allow the extraction of the plasmids incorporated according to the so-called "DNA minipreparation" procedure (17). The plasmid preparation of each recombinant colony was cut with the restriction enzyme *EcoRI* and analyzed on agarose gel. The plasmids possessing an insert detected under UV light after staining the gel with ethidium bromide were selected for the sequencing of the insert after hybridization with a primer complementary to the T7 promoter present on the cloning plasmid of the TA cloning kit[®]. The reaction prior to the sequencing was then carried out according to the method recommended for using the sequencing kit "PRISMTM Ready Reaction Amplitaq[®] FS, DyeDeoxyTM Terminator" (Applied Biosystems, ref. 402119) and the automated sequencing was carried out on the Applied Biosystems 373 A and 377 apparatus, according to the manufacturer's instructions.

The LB13 clone obtained contains an N-terminal region of MSRV-1 gag gene homologous to clone 2 and an

LTR corresponding to a portion of the U5 region. Between the U5 region and gag, a binding site for the transfer RNAs, the PBS "primer binding site", was identified.

5 The nucleotide sequence of the 764 bp fragment of the LB13 clone in the plasmid "pCRTM vector" is represented in the identifier SEQ ID NO: 141.

 The binding site for the transfer RNAs, having a sequence of PBS tryptophan type, was identified at
10 nucleotide position No. 342-359 of the LB13 clone.

 As this same PBS was found in the endogenous copies homologous to MSRV1, the endogenous family thus defined is henceforth called HERV W, according to the nomenclature proposed for the endogenous retrovirus
15 families (W=tryptophan).

 A short ORF of about 65 amino acids was found in the U5 region of the 5' LTR of the LB13 clone.

 Sequence of the ORF:

PMASNRAITLTAWSKIPFLGIRETKNPRSENTRLATMLEAAHHHFGSSPPLSWEL
20 WEQGPQVTIW.

 The corresponding nucleotide sequence starting at an ATG codon is capable of being expressed in a subgenomic DNA from a proviral LTR (U3RU5).

 Another clone, called LA15, was obtained on the
25 total RNA extracted from virions concentrated by ultracentrifugation from a culture supernatant of synoviocytes obtained from a patient suffering from rheumatoid arthritis. The strategy for amplifying and cloning the LA15 clone is exactly the same which was
30 used for the LB13 clone.

 The nucleotide sequence of the LA15 clone, which is represented in the identifier SEQ ID NO: 142, is very similar to the LB13 clone. This suggests that the MSVR-1 retrovirus detected in multiple sclerosis
35 has sequences which are similar to those found in rheumatoid arthritis.

EXAMPLE 8: RECONSTRUCTION OF AN RU5-GAG REGION FROM THE CLONES LB15, LB13, CL2 AND CL17

The clones CL2 and LB13 have already been described in the preceding examples. The LB15 clone was obtained using the R sequence of the LTR of the cl6 clone in order to define a primer in 5' and the anti-sense primers used are the same as for the LB13 clone. The CL17 clone was obtained by nested RT-PCR using the following primers:

5'-TCATGCAACTGCACTCTTCTGGTCCG-3' (sense)

5'-TCTTGCACTAACCTCCACTGTCCGTTGG-3' (antisense)

5'-ATCCCCCAGTAACAATTTGGTGACCACG-3' (sense)

5'-TCGGGTCTAAGAGGGTACTTCCTTTGGTAGG-3' (antisense)

The LB15 clone was obtained from virions obtained by culturing MS cells. The LB17 clone was obtained from culturing plasma from an MS patient.

These overlapping clones made it possible to reconstruct an RU5-gag sequence with a potential ORF in the gag gene, as presented in Figure 14.

EXAMPLE 9: PREPARATION OF A CLONE 87-23

The region corresponding to integrase was amplified and cloned from MS plasma using a seminested RT PCR with the following primers situated in the pol and env regions of MSRV1.

In the pol region:

5'-TTACGCAGGTCTCAGGGATGAGCTT-3' (sense-primary PCR)

5'-CGGCAGTAGCAGTCTTAGTATCTGAAGCAGTTA-3' (sense-secondary PCR)

In the env region,

5'-GGTACGGAGGGTTTCATGTAGTTTTGAG-3' (anti-sense primary and secondary PCR)

The amplified clone contains 774 bp in the pol/RT region, all the integrase region (1197 bp) and

the start of the env region (480 bp). The nucleotide sequence corresponding to the integrase region and the translation to amino acids of the potential ORF are presented in Figure 15.

5

EXAMPLE 10: CONFIRMATION OF THE PRESENCE OF RNA CONTAINING ENV SEQUENCES RELATED TO ERV9 IN THE RETROVIRAL PARTICLES ASSOCIATED WITH THE MSRV1 GENOME:

Sequences related to ERV9 have been found in a
10 minor proportion in the virion preparations obtained from MS compared with the MSRV1 sequences. The existence of phenomena of co-encapsulation of phylogenetically related endogenous sequences into retroviral particles produced by a replicative strain
15 has been described. Surprisingly, an RNA region comprising an ORF starting in the 3' portion of env and continuing potentially into the 3' LTR has been found in various MS samples. In order to specify the existence of an ORF, transcription-translation tests
20 were carried out and made it possible to show the reality of an env ORF containing the entire transmembrane (TM) portion and ending at the start of the putative LTR. However, an additional frame (ORFX) follows and continues in the 3' LTR. The two products
25 of expression were visualized and their respective ORFs were subcloned. Figure 16 represents the nucleotide and peptide sequences of the B13 clone already described, specifying the ORFs in the truncated env region and in the putative LTR. The presence of such RNAs may be
30 responsible for recombinations with the replicative strain and consequently generate strains having a modified pathogenicity.

BIBLIOGRAPHIC REFERENCES

- (1) Laemmli U.K. Cleavage of structural proteins
5 during the assembly of the head of bacteriophage T4.
Nature. (1970). 227: 680-685.
- (2) Towbin H., Staehelin T. & Gordon J.
Electrophoretic transfer of proteins from
10 polyacryalmide gels to nitrocellulose sheets: procedure
and some applications. Proc. Natl. Acad. Sci. USA.
(1979). 76: 4350-4354.

CLAIMS

1. Nucleic material, in isolated or purified state, comprising a nucleotide sequence chosen from the group which consists of (i) the sequences
5 SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences complementary to sequences (i); and (iii) the sequences
10 equivalent to sequences (i) or (ii), in particular the sequences having, for every series of 100 contiguous monomers, at least 50%, and preferentially at least 70% homology with sequences (i) or (ii) respectively.
2. Nucleic material, in isolated or purified
15 state, encoding a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with a peptide sequence chosen from the group which consists of SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118,
20 SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137.
3. Retroviral nucleic material, whose pol gene comprises a nucleotide sequence identical or equivalent to a sequence chosen from the group which consists of SEQ ID NO: 112, SEQ ID NO: 124 and their complementary
25 sequences.
4. Retroviral nucleic material, in which the 5' end of the pol gene starts at nucleotide 1419 of SEQ ID NO: 130.
5. Retroviral nucleic material, in which the pol
30 gene encodes a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with the peptide sequence SEQ ID NO: 113.
6. Retroviral nucleic material, in which the 3'
35 end of the gag gene ends at nucleotide 1418 of SEQ ID NO: 130.
7. Retroviral nucleic material, in which the env gene comprises a nucleotide sequence identical or

equivalent to a sequence chosen from the group which consists of SEQ ID NO: 117, and its complementary sequences.

5 8. Retroviral nucleic material, in which the env gene comprises a nucleotide sequence which starts at nucleotide 1 of SEQ ID NO: 117 and ends at nucleotide at nucleotide [sic] 233 of SEQ ID NO: 114.

9. Retroviral nucleic material, in which the env gene encodes a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with the sequence SEQ ID NO: 118.

10. Retroviral nucleic material in which the U3R region of the 3' LTR comprises a nucleotide sequence which ends at nucleotide 617 of SEQ ID NO: 114.

11. Retroviral nucleic material in which the RU5 region of the 5' LTR comprises a nucleotide sequence which starts at nucleotide 755 of SEQ ID NO: 120 and ends at nucleotide 337 of SEQ ID NO: 141 or SEQ ID NO: 142.

12. Retroviral nucleic material comprising a sequence which starts at nucleotide 755 of SEQ ID NO: 120 and which ends at nucleotide 617 of SEQ ID NO: 114.

13. Retroviral nucleic material according to any one of the preceding claims, characterized in that it is associated with at least one autoimmune disease such as multiple sclerosis or rheumatoid arthritis.

14. Nucleotide fragment comprising a nucleotide sequence chosen from the group which consists of (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences complementary to sequences (i); and (iii) the sequences equivalent to sequences (i) or (ii), in particular the sequences having, for every series of 100 contiguous monomers, at least 50%, and preferentially at least 70% homology with sequences (i) or (ii) respectively.

15. Nucleotide fragment according to Claim 14, consisting of a nucleotide sequence chosen from the group which consists of (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, 5 SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142; (ii) the sequences complementary to sequences (i); and (iii) the sequences equivalent to sequences (i) or (ii), in particular the sequences having, for every series of 100 contiguous 10 monomers, at least 50%, and preferentially at least 70% homology with sequences (i) or (ii) respectively.

16. Nucleotide fragment comprising a nucleotide sequence encoding a polypeptide having, for every contiguous series of at least 30 amino acids, at least 15 50%, and preferably at least 70% homology with a peptide sequence chosen from the group which consists of SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137.

17. Nucleotide fragment according to claim 16, 20 consisting of a nucleotide sequence encoding a polypeptide having, for every contiguous series of at least 30 amino acids, at least 50%, and preferably at least 70% homology with a peptide sequence chosen from the group which consists of SEQ ID NO: 113, SEQ ID NO: 25 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137.

18. Nucleic probe for the detection of a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, characterized in that it is capable of 30 hybridizing specifically with any fragment according to any one of claims 14 to 17, belonging to the genome of said retrovirus.

19. Probe according to claim 18, characterized in that it possesses from 10 to 100 nucleotides, 35 preferably from 10 to 30 nucleotides.

20. Primer for the amplification, by polymerization, of an RNA or of a DNA of a retrovirus associated with multiple sclerosis and/or rheumatoid

arthritis, characterized in that it comprises a nucleotide sequence identical or equivalent to at least a portion of the nucleotide sequence of a fragment according to any one of claims 8 to 11, in particular a
5 nucleotide sequence having, for every series of 10 contiguous monomers, at least 50%, preferably at least 70% homology with at least said portion of said fragment.

21. Primer according to claim 20, characterized in
10 that its nucleotide sequence is chosen from
SEQ ID NO: 116, SEQ ID NO: 119, SEQ ID NO: 122,
SEQ ID NO: 123, SEQ ID NO: 126, SEQ ID NO: 127,
SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 132, and
SEQ ID NO: 133.

15 22. RNA or DNA, and in particular replication and/or expression vector, comprising a genomic fragment of the nucleic material according to any one of claims 1 to 7 or a fragment according to any one of claims 14 to 17.

20 23. Peptide encoded by any open reading frame belonging to a nucleotide fragment according to any one of claims 14 to 17, in particular a polypeptide, for example oligopeptide forming or comprising an antigenic determinant recognized by sera of patients infected
25 with the MSRV-1 virus, and/or in whom the MSRV-1 virus has been reactivated.

24. Peptide according to claim 23 comprising a sequence identical, partially or completely, or equivalent to a sequence chosen from SEQ ID NO: 113,
30 SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 and SEQ ID NO: 137.

25. Diagnostic, prophylactic or therapeutic composition, in particular for inhibiting the expression of at least one retrovirus associated with
35 multiple sclerosis and/or rheumatoid arthritis, comprising a nucleotide fragment according to any one of claims 14 to 17.

26. Method for detecting a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis, in a biological sample, characterized in that an RNA and/or a DNA assumed to belong to or obtained from said
5 retrovirus, or their complementary RNA and/or DNA, is brought into contact with a composition comprising a nucleotide fragment according to any one of claims 14 to 17.

SEQUENCE LISTING

- (2) INFORMATION FOR SEQ ID NO: 68:
- (i) SEQUENCE CHARACTERISTICS:
- 5 (A) LENGTH: 34 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:
- GACTCGCTGC AGATCGATTT TTTTTTTTTT TTTT 34
- (2) INFORMATION FOR SEQ ID NO: 69:
- (i) SEQUENCE CHARACTERISTICS:
- 15 (A) LENGTH: 30 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:
- GCCATCAAGC CACCCAAGAA CTCTTAAGTT 30
- (2) INFORMATION FOR SEQ ID NO: 70:
- (i) SEQUENCE CHARACTERISTICS:
- 25 (A) LENGTH: 30 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- 30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:
- CCAATAGCCA GACCATTATA TACTACTAATT 30
- (2) INFORMATION FOR SEQ ID NO: 112:
- (i) SEQUENCE CHARACTERISTICS:
- 35 (A) LENGTH: 310 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

```
GCTTATAGAA GGACCCCTAG TATGGGGTAA TCCCCTCTGG GAAACCAAGC CCCAGTACTC 60
AGCAGGAAAA ATAGAATAGG AAACCTCACA AGGACATACT TTCCTCCCCT CCAGATGGCT 120
AGCCACTGAG GAAGGAAAAA TACTTTCACC TGCAGCTAAC CAACAGAAAT TACTTAAAC 180
CCTTCACCAA ACCTTCCACT TAGGCATTGA TAGCACCCAT CAGATGGCCA AATTATTATT 240
TACTGGACCA GGCCTTTTCA AAACATATCAA GAAGATAGTC AGGGGCTGTG AAGTGTGCCA 300
AAGAAATAAT 310
```

(2) INFORMATION FOR SEQ ID NO: 113:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 103 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

```
Leu Ile Glu Gly Pro Leu Val Trp Gly Asn Pro Leu Trp Glu Thr Lys
1           5           10           15
Pro Gln Tyr Ser Ala Gly Lys Ile Glu Xaa Glu Thr Ser Gln Gly His
20           25           30
Thr Phe Leu Pro Ser Arg Trp Leu Ala Thr Glu Glu Gly Lys Ile Leu
35           40           45
Ser Pro Ala Ala Asn Gln Gln Lys Leu Leu Lys Thr Leu His Gln Thr
50           55           60
Phe His Leu Gly Ile Asp Ser Thr His Gln Met Ala Lys Leu Leu Phe
65           70           75           80
Thr Gly Pro Gly Leu Phe Lys Thr Ile Lys Lys Ile Val Arg Gly Cys
85           90           95
Glu Val Cys Gln Arg Asn Asn
100
```

(2) INFORMATION FOR SEQ ID NO: 114:

15 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 635 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

```

CCCTGTATCT TTAACCTCCT TGTAAAGTTT GTCTCTTCCA GAATCAAAAC TGTA AAACTA 60
CAAATTGTTT TTCAAATGGA GCACCAGATG GAGTCCATGA CTAAGATCCA CCGTGGACCC 120
CTGGACCCGGC CTGCTAGCCC ATGCTCCGAT GTTAATGACA TTGAAGGCAC CCCTCCCGAG 180
GAAATGTCAA CTGCACAACC CCTACTATGC CCCAATTCAG CGGGAAGCAG TTAGAGCCGT 240
CATCAGCCAA CCTCCCCAAC AGCACTTGGG TTTTCTGTT GAGAGGGGGG ACTGAGAGAC 300
AGGACTAGCT GGATTTCCTA GGCCAACGAA GAATCCCTAA GCCTAGCTGG GAAGGTGACT 360
GCATCCACCT CTAAACATGG GGCTTGCAAC TTAGCTCACA CCGGACCAAT CAGAGAGCTC 420
ACTAAAATGC TAATTAGGCA AAAATAGGAG GTAAAGAAAT AGCCAATCAT CTATTGCCTG 480
AGAGCACAGC GGGAGGGACA AGGATCGGGA TATAACCCA GGCATTGAG CCGGCAACGG 540
CAACCCCTT TGGTCCCT CCCTTTGTAT GGGCGCTCTG TTTTCACTCT ATTTCACTCT 600
ATTAAATCTT GCAACTGAAA AAAAAAAAAA AAAAA 635

```

(2) INFORMATION FOR SEQ ID NO: 115

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 77 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

```

Pro Cys Ile Phe Asn Leu Leu Val Lys Phe Val Ser Ser Arg Ile Lys
1           5           10           15
Thr Val Lys Leu Gln Ile Val Leu Gln Met Glu His Gln Met Glu Ser
20           25           30
Met Thr Lys Ile His Arg Gly Pro Leu Asp Arg Pro Ala Ser Pro Cys
35           40           45
Ser Asp Val Asn Asp Ile Glu Gly Thr Pro Pro Glu Glu Ile Ser Thr
50           55           60
Ala Gln Pro Leu Leu Cys Pro Asn Ser Ala Gly Ser Ser
65           70           75

```

(2) INFORMATION FOR SEQ ID NO: 116:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

TGGGGTTCCA TTTGTAAGAC CATCTGTAGC TT

32

(2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1481 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

ATGGCCCTCC CTTATCATAC TTTCTCTTT ACTGTTCTCT TACCCCTTT CGCTCTCACT 60
GCACCCCTC CATGCTGCTG TACAACCACT AGCTCCCTT ACCAAGAGTT TCTATGAAGA 120
ACGCGGCTTC CTGGAAATAT TGATGCCCCA TCATATAGGA GTTTATCTAA GGGAAACTCC 180
ACCTTCACTG CCCACACCCA TATGCCCCGC AACTGCTATA ACTCTGCCAC TCTTTGCATG 240
CATGCAATA CTCATTATTG GACAGGGAAA ATGATTATC CTAGTTGTCC TGGAGGACTT 300
GGAGCCACTG TCTGTTGGAC TTAATTCACT CATACCAGTA TGTCTGATGG GGGTGGAAAT 360
CAAGGTCAGG CAAGAGAAAA ACAAGTAAAG GAAGCAATCT CCCAACTGAC CCGGGGACAT 420
AGCACCCCTA GCCCCTACAA AGGACTAGTT CTCTCAAAAC TACATGAAAC CCTCCGTACC 480
CATACTCGCC TGGTGAGCCT ATTTAATACC ACCCTCACTC GGCTCCATGA GGTCTCAGCC 540
CAAAACCCCTA CTAAGTCTTG GATGTGCCTC CCCCTGCACT TCAGGCCATA CATTTCATC 600
CCTGTTCTCTG AACAAATGAA CAACTTCAGC ACAGAAATAA ACACCACTTC CGTTTGTAGTA 660
GGACCTCTTG TTTCCAATCT GGAAATAACC CATACCTCAA ACCTCACCTG TGTAAATTTT 720
AGCAATACTA TAGACACAAC CAGCTCCCAA TGCATCAGGT GGGTAACACC TCCCACACGA 780
ATAGTCTGCC TACCCTCAGG AATATTTTTT GTCTGTGGTA CCTCAGCCTA TCATTGTTTG 840
AATGGCTCTT CAGAATCTAT GTGCTTCCTC TCATTCTTAG TGCCCCCTAT GACCATCTAC 900
ACTGAACAAG ATTTATACAA TCATGTCGTA CCTAAGCCCC ACAACAAAAG AGTACCCATT 960
CTTCCTTTTG TTATCAGAGC AGGAGTGCTA GGCAGACTAG GTACTGGCAT TGGCAGTATC 1020

ACAACCTCTA CTCAGTTCTA CTACAAACTA TCTCAAGAAA TAAATGGTGA CATGGAACAG 1080
 GTCACCTGACT CCCTGGTCAC CTTGCAAGAT CAACTTAACT CCCTAGCAGC AGTAGTCCTT 1140
 CAAAATCGAA GAGCTTTACA CTTGCTAACC GCCAAAAGAG GGGGAACCTG TTTATTTTAA 1200
 GGAGAAGAAC GCTGTTATTA TGTTAATCAA TCCAGAATTG TCACTGAGAA AGTTAAAGAA 1260
 ATTCGAGATC GAATACAATG TAGAGCAGAG GAGCTTCAAA ACACCGAACG CTGGGGCCTC 1320
 CTCAGCCAAT GGATGCCCTG GGTTCCTCCC TTCTTAGGAC CTCTAGCAGC TCTAATATTG 1380
 TTAATCCTCT TTGGACCCTG TATCTTTAAC CTCCTTGTTA AGTTTGTCTC TTCCAGAATT 1440
 GAAGCTGTAA AGCTACAGAT GGTCTTACAA ATGGAACCCC A 1481

(2) INFORMATION FOR SEQ ID NO: 118:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 493 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

Met Ala Leu Pro Tyr His Thr Phe Leu Phe Thr Val Leu Leu Pro Pro
 1 5 10 15
 Phe Ala Leu Thr Ala Pro Pro Pro Cys Cys Cys Thr Thr Ser Ser Ser
 20 25 30
 Pro Tyr Gln Glu Phe Leu Xaa Arg Thr Arg Leu Pro Gly Asn Ile Asp
 35 40 45
 Ala Pro Ser Tyr Arg Ser Leu Ser Lys Gly Asn Ser Thr Phe Thr Ala
 50 55 60
 His Thr His Met Pro Arg Asn Cys Tyr Asn Ser Ala Thr Leu Cys Met
 65 70 75 80
 His Ala Asn Thr His Tyr Trp Thr Gly Lys Met Ile Asn Pro Ser Cys
 85 90 95
 Pro Gly Gly Leu Gly Ala Thr Val Cys Trp Thr Tyr Phe Thr His Thr
 100 105 110
 Ser Met Ser Asp Gly Gly Gly Ile Gln Gly Gln Ala Arg Glu Lys Gln
 115 120 125
 Val Lys Glu Ala Ile Ser Gln Leu Thr Arg Gly His Ser Thr Pro Ser
 130 135 140
 Pro Tyr Lys Gly Leu Val Leu Ser Lys Leu His Glu Thr Leu Arg Thr

145	150	155	160
His Thr Arg Leu Val Ser Leu Phe Asn Thr Thr Leu Thr Arg Leu His			
	165	170	175
Glu Val Ser Ala Gln Asn Pro Thr Asn Cys Trp Met Cys Leu Pro Leu			
	180	185	190
His Phe Arg Pro Tyr Ile Ser Ile Pro Val Pro Glu Gln Trp Asn Asn			
	195	200	205
Phe Ser Thr Glu Ile Asn Thr Thr Ser Val Leu Val Gly Pro Leu Val			
	210	215	220
Ser Asn Leu Glu Ile Thr His Thr Ser Asn Leu Thr Cys Val Lys Phe			
225	230	235	240
Ser Asn Thr Ile Asp Thr Thr Ser Ser Gln Cys Ile Arg Trp Val Thr			
	245	250	255
Pro Pro Thr Arg Ile Val Cys Leu Pro Ser Gly Ile Phe Phe Val Cys			
	260	265	270
Gly Thr Ser Ala Tyr His Cys Leu Asn Gly Ser Ser Glu Ser Met Cys			
	275	280	285
Phe Leu Ser Phe Leu Val Pro Pro Met Thr Ile Tyr Thr Glu Gln Asp			
	290	295	300
Leu Tyr Asn His Val Val Pro Lys Pro His Asn Lys Arg Val Pro Ile			
305	310	315	320
Leu Pro Phe Val Ile Arg Ala Gly Val Leu Gly Arg Leu Gly Thr Gly			
	325	330	335
Ile Gly Ser Ile Thr Thr Ser Thr Gln Phe Tyr Tyr Lys Leu Ser Gln			
	340	345	350
Glu Ile Asn Gly Asp Met Glu Gln Val Thr Asp Ser Leu Val Thr Leu			
	355	360	365
Gln Asp Gln Leu Asn Ser Leu Ala Ala Val Val Leu Gln Asn Arg Arg			
	370	375	380
Ala Leu Asp Leu Leu Thr Ala Lys Arg Gly Gly Thr Cys Leu Phe Leu			
385	390	395	400
Gly Glu Glu Arg Cys Tyr Tyr Val Asn Gln Ser Arg Ile Val Thr Glu			
	405	410	415
Lys Val Lys Glu Ile Arg Asp Arg Ile Gln Cys Arg Ala Glu Glu Leu			
	420	425	430
Gln Asn Thr Glu Arg Trp Gly Leu Leu Ser Gln Trp Met Pro Trp Val			

Leu Pro Phe Leu Gly Pro Leu Ala Ala Leu Ile Leu Leu Leu Leu Phe
450 455 460
Gly Pro Cys Ile Phe Asn Leu Leu Val Lys Phe Val Ser Ser Arg Ile
465 470 475 480
Glu Ala Val Lys Leu Gln Met Val Leu Gln Met Glu Pro
485 490

- (2) INFORMATION FOR SEQ ID NO: 119:
- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:
- TCAAAATCGA AGAGCTTTAG ACTTGCTAAC CG

- ```

(2) INFORMATION FOR SEQ ID NO: 120:
 (i) SEQUENCE CHARACTERISTICS:
15 (A) LENGTH: 1329 base pairs
 (B) TYPE: nucleotide
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: cDNA
20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:
TCAAAATCGA AGAGCTTTAG ACTTGCTAAC CGCCAAAAGA GGGGGAACCT GTTTATTTTT 60
AGGGGAAGAA TGCTGTTAGT ATGTTAATCA ATCTGGAATC ATTACTGAGA AAGTTAAAGA 120
AATTTGAGAT CGAATATAAT GTAGAGCAGA GGACCTTCAA AACACTGCAC CCTGGGGCCT 180
CCTCAGCCAA TGGATGCCCT GGACTCTCCC CTTCTTAGGA CCTCTAGCAG CTATAATATT 240
TTTACTCCTC TTTGGACCCCT GTATCTTCAA CTTCTTGTT AAGTTTGTCT CTTCCAGAAT 300
TGAAGCTGTA AAGCTACAAA TAGTTCCTCA AATGGAACCC CAGATGCAGT CCATGACTAA 360
AATCTACCGT GGACCCCTGG ACCGGCCTGC TAGACTATGC TCTGATGTTA ATGACATTGA 420
AGTCACCCCT CCCGAGGAAA TCTCAACTGC ACAACCCCTA CTACACTCCA ATTCAGTAGG 480
AAGCAGTTAG AGCAGTTGTC AGCCAACCTC CCCAACAGTA CTTGGGTTTT CCTGTTGAGA 540
GGGTGGACTG AGAGACAGGA CTAGCTGGAT TTCCTAGGCT GACTAAGAAT CCCNAAGCCT 600

```

ANCTGGGAAG GTGACCGCAT CCATCTTTAA ACATGGGGCT TGCAACTTAG CTCACACCCG 660  
 ACCAATCAGA GAGCTCACTA AAATGCTAAT CAGGCAAAAA CAGGAGGTAA AGCAATAGCC 720  
 AATCATCTAT TGCCTGAGAG CACAGCGGGA AGGACAAGGA TTGGGATATA AACTCAGGCA 780  
 TTCAAGCCAG CAACAGCAAC CCCCTTTGGG TCCCTCCCA TTGTATGGGA GCTCTGTTTT 840  
 CACTCTATTT CACTCTATTA AATCATGCAA CTGCACTCTT CTGGTCCGTG TTTTTTATGG 900  
 CTCAAGCTGA GCTTTTGTTC GCCATCCACC ACTGCTGTTT GCCACCGTCA CAGACCCGCT 960  
 GCTGACTTCC ATCCCTTTGG ATCCAGCAGA GTGTCCACTG TGCTCCTGAT CCAGCGAGGT 1020  
 ACCCATTGCC ACTCCCGATC AGGCTAAAGG CTGCCATTG TTCCTGCATG GCTAAGTGCC 1080  
 TGGGTTTGTG CTAATAGAAC TGAACACTGG TCACTGGCTT CCATGGTTCT CTTCCATGAC 1140  
 CCACGGCTTC TAATAGAGCT ATAACACTCA CCGCATGGCC CAAGATTCCA TTCCTTGCTA 1200  
 TCTGTGAGGC CAAGAACCCC AGGTCAGAGA ANGTCAGGCT TGCCACCATT TGGGAAGTGG 1260  
 CCCACTGCCA TTTTGCTAGC GGCCACCCAC CATCTTGGGA GCTGTGGGAG CAAGGATCCC 1320  
 CCAGTAACA 1329

(2) INFORMATION FOR SEQ ID NO: 121:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 162 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Gln | Asn | Arg | Arg | Ala | Leu | Asp | Leu | Leu | Thr | Ala | Lys | Arg | Gly | Gly | Thr |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Cys | Leu | Phe | Leu | Gly | Glu | Glu | Cys | Cys | Xaa | Tyr | Val | Asn | Gln | Ser | Gly |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Ile | Ile | Thr | Glu | Lys | Val | Lys | Glu | Ile | Xaa | Asp | Arg | Ile | Xaa | Cys | Arg |
|     |     |     | 35  |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| Ala | Glu | Asp | Leu | Gln | Asn | Thr | Ala | Pro | Trp | Gly | Leu | Leu | Ser | Gln | Trp |
|     |     |     | 50  |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| Met | Pro | Trp | Thr | Leu | Pro | Phe | Leu | Gly | Pro | Leu | Ala | Ala | Ile | Ile | Phe |
| 65  |     |     |     | 70  |     |     |     | 75  |     |     |     |     | 80  |     |     |
| Leu | Leu | Leu | Phe | Gly | Pro | Cys | Ile | Phe | Asn | Phe | Leu | Val | Lys | Phe | Val |
|     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |     |
| Ser | Ser | Arg | Ile | Glu | Ala | Val | Lys | Leu | Gln | Ile | Val | Leu | Gln | Met | Glu |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |

Pro Gln Met Gln Ser Met Thr Lys Ile Tyr Arg Gly Pro Leu Asp Arg  
 115 120 125  
 Pro Ala Arg Leu Cys Ser Asp Val Asn Asp Ile Glu Val Thr Pro Pro  
 130 135 140  
 Glu Glu Ile Ser Thr Ala Gln Pro Leu Leu His Ser Asn Ser Val Gly  
 145 150 155 160  
 Ser Ser

(2) INFORMATION FOR SEQ ID NO: 122:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

10 GGCATTGATA GCACCCATCA G 21

(2) INFORMATION FOR SEQ ID NO: 123:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

20 CATGTCACCA GGGTGAATA G 21

(2) INFORMATION FOR SEQ ID NO: 124:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

(2) INFORMATION FOR SEQ ID NO: 124:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 758 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

```
GGCATTGATA GCACCCATCA GATGGCCAAA TCATTATTTA CTGGACCAGG CCTTTTCAAA 60
ACTATCAAGC AGATAGGGCC CGTGAAGCAT GCCAAAGAAA TAATCCCCTG CCTTATCGCC 120
ATGTTCTTTC AGGAGAACAA AGAACAGGCC ATTACCCAGG GGAAGACTGG CAACTAGATT 180
TTACCCACAT GGCCAAATGT CAGGGATTTC AGCATCTACT AGTCTGGGCA GATACTTTCA 240
CTGGTTGGGT GGAGTCTTCT CCTTGTAGGA CAGAAAAGAC CCAAGAGGTA ATAAAGGCAC 300
TAATGAAATA ATTCCCAGAT TTGGACTTCC CCCAGGATTA CAGGGTGACA ATGCCCCCGC 360
TTTCAAGGCT GCAGTAACCC AGGGAGTATC CCAGGTGTTA GGCATACAAT ATCACTTACA 420
CTGTGCCTGG AGGCCACAAT CCTCCAGAAA AGTCAAGAAA ATGAATGAAA CACTCAAAGA 480
TCTAAAAAAG CTAACCCAAG AAACCCACAT TGCATGACCT GTTCTGTTGC CTATAACCTT 540
ACTAAGAATC CATAACTATC CCCCCAAAAG CAGGACTTAG CCCATACGAG ATGCTATATG 600
GATGGCCTTT CCTAACCAAT GACCTTGTGC TTGACTGAGA AATGGCCAAC TTAGTTGCAG 660
ACATCACCTC CTTAGCCAAA TATCAACAAG TTCTTAAAC ATCACAGGGA ACCTGTCCCC 720
GAGAGGAGGG AAAGGAACTA TTCCACCCTG GTGACATG 758
```

10 (2) INFORMATION FOR SEQ ID NO: 126:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

CGGACATCCA AAGTGATGGG AAACG 25

20 (2) INFORMATION FOR SEQ ID NO: 127:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleotide
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

GGACAGGAAA GTAAGACTGA GAAGGC

26

(2) INFORMATION FOR SEQ ID NO: 128:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 26 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

CCTAGAACGT ATTCTGGAGA ATTGGG

26

(2) INFORMATION FOR SEQ ID NO: 129:

(i) SEQUENCE CHARACTERISTICS:

- 15 (A) LENGTH: 26 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

TGGCTCTCAA TGGTCAAACA TACCCG

26

(2) INFORMATION FOR SEQ ID NO: 130:

(i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 1511 base pairs  
(B) TYPE: nucleotide  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

CCTAGAACGT ATTCTGGAGA ATTGGGACCA ATGTGACACT CAGACGCTAA GAAAGAAACG 60

```

ATTATATTC TTCTGCAGTA CCGCCTGGCC ACAATATCCT CTTCAAGGGA GACAAACCTG 120
GCTTCCTGAG GGAAGTATAA ATTATAACAT CATCTTACAG CTAGACCTCT TCTGTAGAAA 180
GGAGGGCAAA TGGAGTGAAG TGCCATATGT GCAAACTTTC TTTTCATTAA GAGACAACTC 240
ACAATTATGT AAAAAGTGTG GTTTATGCCC TACAGGAAGC CCTCAGAGTC CACCTCCCTA 300
CCCCAGCGTC CCCTCCCCGA CTCCTTCCTC AACTAATAAG GACCCCCCTT TAACCCAAAC 360
GGTCCAAAAG GAGATAGACA AAGGGGTAA CAATGAACCA AAGAGTGCCA ATATTCCCCG 420
ATTATGCCCC CTCCAAGCAG TCAGAGGAGG AGAATTGCGC CCAGCCAGAG TGCCTGTACC 480
TTTTTCTCTC TCAGACTTAA AGCAAATTAA AATAGACCTA GGTAATTCT CAGATAACCC 540
TGACCGCTAT ATTGATGTTT TACAAGGGTT AGGACAATCC TTTGATCTGA CATGGAGAGA 600
TATAATGTTA CTACTAAATC AGACACTAAC CCAAATGAG ACAAGTGCCG CTGTAACTGC 660
AGCCCCGAGAG TTTGGCGATC TTTGGTATCT CAGTCAGGCC AACAATAGGA TGACAACAGA 720
GGAAGAACA ACTCCACAG GCCAGCAGGC AGTTCCCACT GTAGACCTC ATTGGGACAC 780
AGAATCAGAA CATGGAGATT GGTGCCACAA ACATTTGCTA ACTTGCGTGC TAGAAGGACT 840
GAGGAAACT AGGAAGAAGC CTATGAATTA CTCAATGATG TCCACTATAA CACAGGGAAA 900
GGAAGAAAAT CTTACTGCTT TTCTGGACAG ACTAAGGGAG GCATTGAGGA AGCATACCTC 960
CCTGTCACCT GACTCTATTG AAGGCCAACT AATCTTAAAG GATAAGTTA TCACTCAGTC 1020
AGCTGCAGAC ATTAGAAAAA ACTTCAAAAG TCTGCCTTAG GCCCGGAGCA GAACTTAGAA 1080
ACCCTATTTA ACTTGCCATC CTCAGTTTTT TATAATAGAG ATCAGGAGGA GCAGGCGAAA 1140
CGGGACAAAC GGGATAAAAA AAAAAGGGGG GGTCCACTAC TTTAGTCATG GCCCTCAGGC 1200
AAGCAGACTT TGGAGGCTCT GCAAAAGGGA AAAGCTGGGC AAATCAAATG CCTAATAGGG 1260
CTGGCTTCCA GTGCGGTCTA CAAGGACACT TAAAAAAGA TTATCCAAGT AGAAATAAGC 1320
CGCCCCCTTG TCCATGCCCC TTACGTCAAG GGAATCACTG GAAGGCCAC TGCCCCAGGG 1380
GATGAAGATA CTCTGAGTCA GAAGCCATTA ACCAGATGAT CCAGCAGCAG GACTGAGGGT 1440
GCCCCGGGCG AGCGCCAGCC CATGCCATCA CCCTCACAGA GCCCCGGGTA TGTTTGACCA 1500
TTGAGAGCCA A
1511

```

(2) INFORMATION FOR SEQ ID NO: 131:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 352 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

Leu Glu Arg Ile Leu Glu Asn Trp Asp Gln Cys Asp Thr Gln Thr Leu

10

1

5

10

15



|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Arg | Lys | Lys | Arg | Phe | Ile | Phe | Phe | Cys | Ser | Thr | Ala | Trp | Pro | Gln | Tyr |
| 20  |     |     |     | 25  |     |     |     | 30  |     |     |     |     |     |     |     |
| Pro | Leu | Gln | Gly | Arg | Glu | Thr | Trp | Leu | Pro | Glu | Gly | Ser | Ile | Asn | Tyr |
| 35  |     |     |     | 40  |     |     |     | 45  |     |     |     |     |     |     |     |
| Asn | Ile | Ile | Leu | Gln | Leu | Asp | Leu | Phe | Cys | Arg | Lys | Glu | Gly | Lys | Trp |
| 50  |     |     |     | 55  |     |     |     | 60  |     |     |     |     |     |     |     |
| Ser | Glu | Val | Pro | Tyr | Val | Gln | Thr | Phe | Phe | Ser | Leu | Arg | Asp | Asn | Ser |
| 65  |     |     |     | 70  |     |     |     | 75  |     |     |     | 80  |     |     |     |
| Gln | Leu | Cys | Lys | Lys | Cys | Gly | Leu | Cys | Pro | Thr | Gly | Ser | Pro | Gln | Ser |
| 85  |     |     |     | 90  |     |     |     | 95  |     |     |     |     |     |     |     |
| Pro | Pro | Pro | Tyr | Pro | Ser | Val | Pro | Ser | Pro | Thr | Pro | Ser | Ser | Thr | Asn |
| 100 |     |     |     | 105 |     |     |     | 110 |     |     |     |     |     |     |     |
| Lys | Asp | Pro | Pro | Leu | Thr | Gln | Thr | Val | Gln | Lys | Glu | Ile | Asp | Lys | Gly |
| 115 |     |     |     | 120 |     |     |     | 125 |     |     |     |     |     |     |     |
| Val | Asn | Asn | Glu | Pro | Lys | Ser | Ala | Asn | Ile | Pro | Arg | Leu | Cys | Pro | Leu |
| 130 |     |     |     | 135 |     |     |     | 140 |     |     |     |     |     |     |     |
| Gln | Ala | Val | Arg | Gly | Gly | Glu | Phe | Gly | Pro | Ala | Arg | Val | Pro | Val | Pro |
| 145 |     |     |     | 150 |     |     |     | 155 |     |     |     | 160 |     |     |     |
| Phe | Ser | Leu | Ser | Asp | Leu | Lys | Gln | Ile | Lys | Ile | Asp | Leu | Gly | Lys | Phe |
| 165 |     |     |     | 170 |     |     |     | 175 |     |     |     |     |     |     |     |
| Ser | Asp | Asn | Pro | Asp | Gly | Tyr | Ile | Asp | Val | Leu | Gln | Gly | Leu | Gly | Gln |
| 180 |     |     |     | 185 |     |     |     | 190 |     |     |     |     |     |     |     |
| Ser | Phe | Asp | Leu | Thr | Trp | Arg | Asp | Ile | Met | Leu | Leu | Leu | Asn | Gln | Thr |
| 195 |     |     |     | 200 |     |     |     | 205 |     |     |     |     |     |     |     |
| Leu | Thr | Pro | Asn | Glu | Arg | Ser | Ala | Ala | Val | Thr | Ala | Ala | Arg | Glu | Phe |
| 210 |     |     |     | 215 |     |     |     | 220 |     |     |     |     |     |     |     |
| Gly | Asp | Leu | Trp | Tyr | Leu | Ser | Gln | Ala | Asn | Asn | Arg | Met | Thr | Thr | Glu |
| 225 |     |     |     | 230 |     |     |     | 235 |     |     |     | 240 |     |     |     |
| Glu | Arg | Thr | Thr | Pro | Thr | Gly | Gln | Gln | Ala | Val | Pro | Ser | Val | Asp | Pro |
| 245 |     |     |     | 250 |     |     |     | 255 |     |     |     |     |     |     |     |
| His | Trp | Asp | Thr | Glu | Ser | Glu | His | Gly | Asp | Trp | Cys | His | Lys | His | Leu |
| 260 |     |     |     | 265 |     |     |     | 270 |     |     |     |     |     |     |     |
| Leu | Thr | Cys | Val | Leu | Glu | Gly | Leu | Arg | Lys | Thr | Arg | Lys | Lys | Pro | Met |
| 275 |     |     |     | 280 |     |     |     | 285 |     |     |     |     |     |     |     |
| Asn | Tyr | Ser | Met | Met | Ser | Thr | Ile | Thr | Gln | Gly | Lys | Glu | Glu | Asn | Leu |
| 290 |     |     |     | 295 |     |     |     | 300 |     |     |     |     |     |     |     |

- 14 -

|   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|   | Thr | Ala | Phe | Leu | Asp | Arg | Leu | Arg | Glu | Ala | Leu | Arg | Lys | His | Thr | Ser |
|   | 305 |     |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     | 320 |
|   | Leu | Ser | Pro | Asp | Ser | Ile | Glu | Gly | Gln | Leu | Ile | Leu | Lys | Asp | Lys | Phe |
|   |     |     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |
| 5 | Ile | Thr | Gln | Ser | Ala | Ala | Asp | Ile | Arg | Lys | Asn | Phe | Lys | Ser | Leu | Pro |
|   |     |     |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |

## (2) INFORMATION FOR SEQ ID NO: 132:

## (i) SEQUENCE CHARACTERISTICS:

- 10 (A) LENGTH: 30 base pairs  
 (B) TYPE: nucleotide  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

TGCTGGAATT CGGGATCCTA GAACGTATTC

30

## (2) INFORMATION FOR SEQ ID NO: 133:

## (i) SEQUENCE CHARACTERISTICS:

- 20 (A) LENGTH: 30 base pairs  
 (B) TYPE: nucleotide  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

AGTTCTGCTC CGAAGCTTAG GCAGACTTTT

30

## (2) INFORMATION FOR SEQ ID NO: 135:

## (i) SEQUENCE CHARACTERISTICS:

- 30 (A) LENGTH: 398 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro

|                                                                 |     |     |     |
|-----------------------------------------------------------------|-----|-----|-----|
| 1                                                               | 5   | 10  | 15  |
| Arg Gly Ser His Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg |     |     |     |
| 20                                                              | 25  | 30  |     |
| Ile Leu Glu Arg Ile Leu Glu Asn Trp Asp Gln Cys Asp Thr Gln Thr |     |     |     |
| 35                                                              | 40  | 45  |     |
| Leu Arg Lys Lys Arg Phe Ile Phe Phe Cys Ser Thr Ala Trp Pro Gln |     |     |     |
| 50                                                              | 55  | 60  |     |
| Tyr Pro Leu Gln Gly Arg Glu Thr Trp Leu Pro Glu Gly Ser Ile Asn |     |     |     |
| 65                                                              | 70  | 75  | 80  |
| Tyr Asn Ile Ile Leu Gln Leu Asp Leu Phe Cys Arg Lys Glu Gly Lys |     |     |     |
| 85                                                              | 90  | 95  |     |
| Trp Ser Glu Val Pro Tyr Val Gln Thr Phe Phe Ser Leu Arg Asp Asn |     |     |     |
| 100                                                             | 105 | 110 |     |
| Ser Gln Leu Cys Lys Lys Cys Gly Leu Cys Pro Thr Gly Ser Pro Gln |     |     |     |
| 115                                                             | 120 | 125 |     |
| Ser Pro Pro Pro Tyr Pro Ser Val Pro Ser Pro Thr Pro Ser Ser Thr |     |     |     |
| 130                                                             | 135 | 140 |     |
| Asn Lys Asp Pro Pro Leu Thr Gln Thr Val Gln Lys Glu Ile Asp Lys |     |     |     |
| 145                                                             | 150 | 155 | 160 |
| Gly Val Asn Asn Glu Pro Lys Ser Ala Asn Ile Pro Arg Leu Cys Pro |     |     |     |
| 165                                                             | 170 | 175 |     |
| Leu Gln Ala Val Arg Gly Gly Glu Phe Gly Pro Ala Arg Val Pro Val |     |     |     |
| 180                                                             | 185 | 190 |     |
| Pro Phe Ser Leu Ser Asp Leu Lys Gln Ile Lys Ile Asp Leu Gly Lys |     |     |     |
| 195                                                             | 200 | 205 |     |
| Phe Ser Asp Asn Pro Asp Gly Tyr Ile Asp Val Leu Gln Gly Leu Gly |     |     |     |
| 210                                                             | 215 | 220 |     |
| Gln Ser Phe Asp Leu Thr Trp Arg Asp Ile Met Leu Leu Leu Asn Gln |     |     |     |
| 225                                                             | 230 | 235 | 240 |
| Thr Leu Thr Pro Asn Glu Arg Ser Ala Ala Val Thr Ala Ala Arg Glu |     |     |     |
| 245                                                             | 250 | 255 |     |
| Phe Gly Asp Leu Trp Tyr Leu Ser Gln Ala Asn Asn Arg Met Thr Thr |     |     |     |
| 260                                                             | 265 | 270 |     |
| Glu Glu Arg Thr Thr Pro Thr Gly Gln Gln Ala Val Pro Ser Val Asp |     |     |     |
| 275                                                             | 280 | 285 |     |
| Pro His Trp Asp Thr Glu Ser Glu His Gly Asp Trp Cys His Lys His |     |     |     |

|                                                                 |                                             |         |
|-----------------------------------------------------------------|---------------------------------------------|---------|
| 290                                                             | 295                                         | 300     |
| Leu Leu Thr Cys Val                                             | Leu Glu Gly Leu Arg Lys Thr Arg Lys Lys Pro |         |
| 305                                                             | 310                                         | 315 320 |
| Met Asn Tyr Ser Met Met Ser Thr Ile Thr Gln Gly Lys Glu Glu Asn |                                             |         |
| 325                                                             | 330                                         | 335     |
| Leu Thr Ala Phe Leu Asp Arg Leu Arg Glu Ala Leu Arg Lys His Thr |                                             |         |
| 340                                                             | 345                                         | 350     |
| Ser Leu Ser Pro Asp Ser Ile Glu Gly Gln Leu Ile Leu Lys Asp Lys |                                             |         |
| 355                                                             | 360                                         | 365     |
| Phe Ile Thr Gln Ser Ala Ala Asp Ile Arg Lys Asn Phe Lys Ser Leu |                                             |         |
| 370                                                             | 375                                         | 380     |
| Pro Lys Leu Ala Ala Ala Leu Glu His His His His His His         |                                             |         |
| 385                                                             | 390                                         | 395     |

(2) INFORMATION FOR SEQ ID NO: 137:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 378 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

|                                                                 |     |       |
|-----------------------------------------------------------------|-----|-------|
| Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Ile Leu Glu Arg |     |       |
| 1                                                               | 5   | 10 15 |
| Ile Leu Glu Asn Trp Asp Gln Cys Asp Thr Gln Thr Leu Arg Lys Lys |     |       |
| 20                                                              | 25  | 30    |
| Arg Phe Ile Phe Phe Cys Ser Thr Ala Trp Pro Gln Tyr Pro Leu Gln |     |       |
| 35                                                              | 40  | 45    |
| Gly Arg Glu Thr Trp Leu Pro Glu Gly Ser Ile Asn Tyr Asn Ile Ile |     |       |
| 50                                                              | 55  | 60    |
| Leu Gln Leu Asp Leu Phe Cys Arg Lys Glu Gly Lys Trp Ser Glu Val |     |       |
| 65                                                              | 70  | 75 80 |
| Pro Tyr Val Gln Thr Phe Phe Ser Leu Arg Asp Asn Ser Gln Leu Cys |     |       |
| 85                                                              | 90  | 95    |
| Lys Lys Cys Gly Leu Cys Pro Thr Gly Ser Pro Gln Ser Pro Pro Pro |     |       |
| 100                                                             | 105 | 110   |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Tyr | Pro | Ser | Val | Pro | Ser | Pro | Thr | Pro | Ser | Ser | Thr | Asn | Lys | Asp | Pro |
|     | 115 |     |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Pro | Leu | Thr | Gln | Thr | Val | Gln | Lys | Glu | Ile | Asp | Lys | Gly | Val | Asn | Asn |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Glu | Pro | Lys | Ser | Ala | Asn | Ile | Pro | Arg | Leu | Cys | Pro | Leu | Gln | Ala | Val |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
| Arg | Gly | Gly | Glu | Phe | Gly | Pro | Ala | Arg | Val | Pro | Val | Pro | Phe | Ser | Leu |
|     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
| Ser | Asp | Leu | Lys | Gln | Ile | Lys | Ile | Asp | Leu | Gly | Lys | Phe | Ser | Asp | Asn |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     |     | 190 |     |
| Pro | Asp | Gly | Tyr | Ile | Asp | Val | Leu | Gln | Gly | Leu | Gly | Gln | Ser | Phe | Asp |
|     | 195 |     |     |     |     | 200 |     |     |     |     |     | 205 |     |     |     |
| Leu | Thr | Trp | Arg | Asp | Ile | Met | Leu | Leu | Leu | Asn | Gln | Thr | Leu | Thr | Pro |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Asn | Glu | Arg | Ser | Ala | Ala | Val | Thr | Ala | Ala | Arg | Glu | Phe | Gly | Asp | Leu |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Trp | Tyr | Leu | Ser | Gln | Ala | Asn | Asn | Arg | Met | Thr | Thr | Glu | Glu | Arg | Thr |
|     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
| Thr | Pro | Thr | Gly | Gln | Gln | Ala | Val | Pro | Ser | Val | Asp | Pro | His | Trp | Asp |
|     |     |     | 260 |     |     |     |     |     | 265 |     |     |     | 270 |     |     |
| Thr | Glu | Ser | Glu | His | Gly | Asp | Trp | Cys | His | Lys | His | Leu | Leu | Thr | Cys |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Val | Leu | Glu | Gly | Leu | Arg | Lys | Thr | Arg | Lys | Lys | Pro | Met | Asn | Tyr | Ser |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Met | Met | Ser | Thr | Ile | Thr | Gln | Gly | Lys | Glu | Glu | Asn | Leu | Thr | Ala | Phe |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| Leu | Asp | Arg | Leu | Arg | Glu | Ala | Leu | Arg | Lys | His | Thr | Ser | Leu | Ser | Pro |
|     |     |     |     | 325 |     |     |     |     | 330 |     |     |     | 335 |     |     |
| Asp | Ser | Ile | Glu | Gly | Gln | Leu | Ile | Leu | Lys | Asp | Lys | Phe | Ile | Thr | Gln |
|     |     | 340 |     |     |     |     |     | 345 |     |     |     | 350 |     |     |     |
| Ser | Ala | Ala | Asp | Ile | Arg | Lys | Asn | Phe | Lys | Ser | Leu | Pro | Lys | Leu | Ala |
|     |     | 355 |     |     |     |     | 360 |     |     |     |     | 365 |     |     |     |
| Ala | Ala | Leu | Glu | His | His | His | His | His | His | His | His | His | His | His | His |
|     | 370 |     |     |     |     |     | 375 |     |     |     |     |     |     |     |     |

(2) INFORMATION FOR SEQ ID NO: 138:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

5

REPLACEMENT SHEET (RULE 26)

- (D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: cDNA  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:  
CTTGGAGGGT GCATAACCAG GGAAT 25
- 5
- (2) INFORMATION FOR SEQ ID NO: 139:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleotide  
10 (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: cDNA  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:  
TGTCCGCTGT GCTCCTGATC 20
- 15
- (2) INFORMATION FOR SEQ ID NO: 140:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 base pairs  
(B) TYPE: nucleotide  
20 (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: cDNA  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:  
CTATGTCCTT TTGGA CTGTT TGGGT 25
- 25
- (2) INFORMATION FOR SEQ ID NO: 141:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 764 base pairs  
(B) TYPE: nucleotide  
30 (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: cDNA  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

|            |            |            |             |            |            |     |
|------------|------------|------------|-------------|------------|------------|-----|
| TGTCGGCTGT | GCTCCTGATC | CAGCACAGGC | GCCCATTTGCC | TCTCCCAATT | GGGCTAAAGG | 60  |
| CTTGCCATTG | TTCTGCACA  | GCTAAGTGCC | TGGGTTTCATC | CTAATCGAGC | TGAACACTAG | 120 |
| TCAGTGGGTT | CCACGGTTCT | CTTCCATGAC | CCATGGCTTC  | TAATAGAGCT | ATAACACTCA | 180 |
| CTGCATGGTC | CAAGATTCCA | TTCTTTGGAA | TCCGTGAGAC  | CAAGAACCCC | AGGTCAGAGA | 240 |
| ACACAAGGCT | TGCCACCATG | TTGGAAGCAG | CCCACCACCA  | TTTTGGAAGC | AGCCCGCCAC | 300 |
| TATCTTGGGA | GCTCTGGGAG | CAAGGACCCC | AGGTAACAAT  | TTGGTGACCA | CGAAGGGACC | 360 |
| TGAATCCGCA | ACCATGAAGG | GATCTCCAAA | GCAATTGGAA  | ATGTTCTCTC | CAAGGCCAAA | 420 |
| ATGCCCCTAA | GATGTATTCT | GGAGAATTGG | GACCAATTTG  | ACCCTCAGAC | AGTAAGAAAA | 480 |
| AAATGACTTA | TATTCTTCTG | CAGTACCGCC | CTGGCCACGA  | TATCTCTTC  | AAGGGGGAGA | 540 |
| AACCTGGCCT | CCTCAGGGAA | GTATAAATTA | TAACACCATC  | TTACAGCTAG | ACCTGTTTTG | 600 |
| TAGAAAAGCA | GGCAAATGGA | GTGAAGTGCC | ATATTACAA   | ACTTTCTTTT | CATTAAAAGA | 660 |
| CAACTCGCAA | TTATGTTAAC | AGTGTGATTT | GTGTTCTTAC  | ACGGAAGCCC | TCAGATTCTA | 720 |
| CTCCCCACCC | CCGGCATCTC | CCCTGAATCC | CTCCCCAACT  | TATT       |            | 764 |

(2) INFORMATION FOR SEQ ID NO: 142:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 800 base pairs

(B) TYPE: nucleotide

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

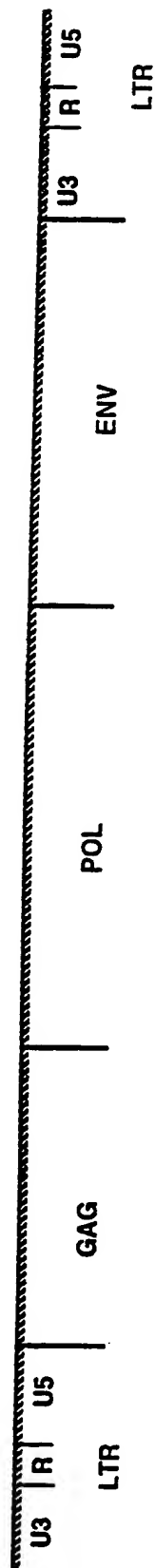
(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

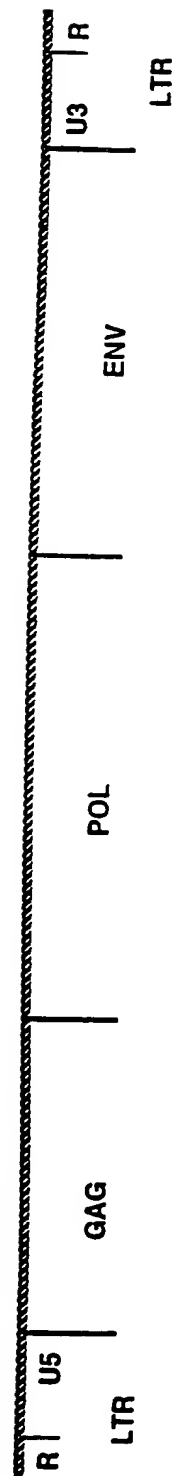
|            |            |            |             |            |             |     |
|------------|------------|------------|-------------|------------|-------------|-----|
| TGTCGGCTGT | GCTCCTGATC | CAGCACAGGC | GCCCATTTGCC | TCTCCCAATT | GGGCTAAAGG  | 60  |
| CTTGCCATTG | TTCTGCACA  | GCTAAGTGCC | TGGGTTTCATC | CTAATCGAGC | TGAACACTAG  | 120 |
| TCAGTGGGTT | CCACGGTTCT | CTTCCATGAC | CCATGGCTTC  | TAATAGAGCT | ATAACACTCA  | 180 |
| CTGCATGGTC | CAAGATTCCA | TTCTTTGGAA | TCCGTGAGAC  | CAAGAACCCC | AGGTCAGAGA  | 240 |
| ACACAAGGCT | TGCCACCATG | TTGGAAGCAG | CCCACCACCA  | TTTTGGAAGC | AGCCCGCCAC  | 300 |
| TATCTTGGGA | GCTCTGGGAG | CAAGGACCCC | CAGGTAACAA  | TTTGGTGACC | ACGAAGGGAC  | 360 |
| CTGAATCCGC | AACCATGAAG | GGATCTCCAA | AGCAATTGGA  | AATGTTCTCT | CCAAGGCCAA  | 420 |
| AATGCCCTTA | AGATGTATT  | TGGAGAATTG | GGACCAATCT  | GACCCTCAGA | CAGTAAGAAA  | 480 |
| AAAAATGACT | TATATTCTTC | TGCAGTACCG | CCTGGCCACG  | GATATCTCT  | TCAAGGGGGA  | 540 |
| GAAACCTGGC | CTCCTGAGGG | AAGTATAAAT | TATAACACCA  | TCTTACAGCT | AGACCTGTTT  | 600 |
| TGTAGAAAAG | GAGGCAAATG | GAGTGAAGTG | CCATATTTAC  | AAACTTTCTT | TTCAATTAATA | 660 |
| GACAACCTGC | AATTATGTAA | ACAGTGTGAT | TTGTGTCCTA  | CAGGAAGCCC | TCAGATCTAC  | 720 |
| CTCCCTACCC | CGGCATCTCC | CTGACTCCTT | CCCCAACTAA  | TAAGGACCCA | CTTCAGCCCA  | 780 |
| AACAGTCCAA | AAGGACATAG |            |             |            |             | 800 |

1 / 32  
FIG 1

PROVIRAL DNA



GENOMIC RNA (VIRION)





2 / 32  
FIG 2

10 20 30 40 50  
1234567890 1234567890 1234567890 1234567890 1234567890  
GCTTATAGAA GGACCCCTAG TATGGGGTAA TCCCTCTGG GAAACCAAGC 50  
A Y R R T P S M G . S P L G N Q A  
L I E G P L V W G N P L W E T K P  
L . K D P . Y G V I P S G K P S  
CCAGTACTC AGCAGGAAAA ATAGAATAGG AAACCTCACA AGGACATACT 100  
P V L S R K N R I G N L T R T Y F  
Q Y S A G K I E . E T S Q G H T  
P S T Q Q E K . N R K P H K D I L  
TTCTTCCCT CCAGATGGCT AGCCACTGAG GAAGGAAAA TACTTTCACC 150  
P P L Q M A S H . G R K N T F T  
F L P S R W L A T E E G K I L S P  
S S P P D G . P L R K E K Y F H L  
TGCAGCTAAC CAACAGAAAT TACTTAAAC CCTTCACCA ACCTTCCACT 200  
C S . P T E I T . N P S P N L P L  
A A N Q Q K L L K T L H Q T F H L  
Q L T N R N Y L K P F T K P S T  
TAGGCATTGA TAGCACCCAT CAGATGGCCA AATTATTATT TACTGGACCA 250  
R H . . H P S D G Q I I I Y W T R  
G I D S T H Q M A K L L F T G P  
. A L I A P I R W P N Y Y L L D Q  
GGCTTTTCA AAACATCAA GAAGATAGTC AGGGGCTGTG AAGTGTGCCA 300  
P F Q N Y Q E D S Q G L . S V P  
G L F K T I K K I V R G C E V C Q  
A F S K L S R R . S G A V K C A K  
AAGAAATAAT  
K K . 310  
R N N  
E I

3 / 32  
FIG 2 (continued)

| 10         | 20          | 30         | 40         | 50         |     |
|------------|-------------|------------|------------|------------|-----|
| 1234567890 | 1234567890  | 1234567890 | 1234567890 | 1234567890 |     |
| CCCTGATCT  | TTAACCTCT   | TGTTAAGTTT | GTCCTTCCA  | GAATCAAAC  | 50  |
| P C I F    | N L L       | V K F      | V S S R    | I K T      |     |
| P V S      | L T S L     | L S L      | S L P      | E S K L    |     |
| L Y L      | . P P       | C . V C    | L F Q      | N Q N      |     |
| TGTAATACTA | CAAATTGTTT  | TTCAAATGGA | GCACCAGATG | GAGTCCATGA | 100 |
| V K L      | Q I V L     | Q M E      | H Q M      | E S M T    |     |
| . N Y      | K L F       | F K W S    | T R W      | S P .      |     |
| C K T T    | N C S       | S N G      | A P D G    | V H D      |     |
| CTAAGATCCA | CCGTGGACCC  | CTGGACCGGC | CTGCTAGCCC | ATGCTCCGAT | 150 |
| K I H      | R G P       | L D R P    | A S P      | C S D      |     |
| L R S T    | V D P       | W T G      | L L A H    | A P M      |     |
| . D P      | P W T P     | G P A C    | . P        | M L R C    |     |
| GTAAATGACA | TTGAAGGCAC  | CCCTCCCGAG | GAAATCTCAA | CTGCACAACC | 200 |
| V N D I    | E G T       | P P E      | E I S T    | A Q P      |     |
| L M T      | L K A P     | L P R      | K S Q      | L H N P    |     |
| . . H      | . R H       | P S R G    | N L N      | C T T      |     |
| CCTACTATGC | CCCAATTTCAG | CGGGAAGCAG | TTAGAGCGGT | CATCAGCCAA | 250 |
| L L C      | P N S A     | G S S      | . S G      | H Q P T    |     |
| Y Y A      | P I Q       | R E A V    | R A V      | I S Q      |     |
| P T M P    | Q F S       | G K Q      | L E R S    | S A N      |     |
| CCTCCCCAAC | AGCACTGGG   | TTTTCTGTT  | GAGAGGGGG  | ACTGAGAGAC | 300 |
| S P T      | A L G       | F S C      | . E G G    | L R D      |     |
| P P Q Q    | H L G       | F P V      | E R G D    | . E T      |     |
| L P N      | S T W V     | F L L      | R G G      | T E R Q    |     |
| AGGACTAGCT | GGATTTCCTA  | GGCCAACGAA | GAATCCCTAA | GCCTAGCTGG | 350 |
| R T S W    | I S .       | A N E      | E S L S    | L A G      |     |
| G L A      | G F P R     | P T K      | N P .      | A . L G    |     |
| D . L      | D F L       | G Q R R    | I P K      | P S W      |     |

4 / 32  
FIG 3

| 10         | 20         | 30          | 40          | 50          |     |
|------------|------------|-------------|-------------|-------------|-----|
| 1234567890 | 1234567890 | 1234567890  | 1234567890  | 1234567890  |     |
| GAAGGTGACT | GCATCCACCT | CTAAACATGG  | GGCTTGCAAC  | TTAGCTCACA  | 400 |
| K V T      | A S T S    | K H G       | A C N       | L A H T     |     |
| R .        | L H P P    | L N M G     | L A T       | . L T       |     |
| E G D C    | I H L .    | T W         | G L Q L     | S S H       |     |
| COCGACCAAT | CAGAGAGCTC | ACTAAAATGC  | TAATTAGGCA  | AAAATAGGAG  | 450 |
| R P I      | R E L      | T K M L     | I R Q       | K . E       |     |
| P D Q S    | E S S      | L K C       | . L G K     | N R R       |     |
| P T N      | Q R A H    | . N A       | N . A       | K I G G     |     |
| GTAAAGAAAT | AGCCAATCAT | CTATTGCCTG  | AGAGCACAGC  | GGGAGGGACA  | 500 |
| V K K .    | P I I      | Y C L       | R A Q R     | E G Q       |     |
| . R N      | S Q S S    | I A .       | E H S       | G R D K     |     |
| K E I      | A N H      | L L P E     | S T A       | G G T       |     |
| AGGATCGGGA | TATAAACCCA | GGCATTGCGAG | COGGCAAACGG | CAACCCCOCTT | 550 |
| G S G      | Y K P R    | H S S       | R Q R       | Q P P L     |     |
| D R D      | I N P      | G I R A     | G N G       | N P L       |     |
| R I G I    | . T Q      | A F E       | P A T A     | T P F       |     |
| TGGGTCCOCT | CCCTTGTAT  | GGCGCTCTG   | TTTCACTCT   | ATTCACTCT   | 600 |
| G P L      | P L Y      | G R S V     | F T L       | F H S       |     |
| W V P S    | L C M      | G A L       | F S L Y     | F T L       |     |
| G S P      | P F V W    | A L C       | F H S       | I S L Y     |     |
| ATTAAATCTT | GCAACTGAAA | AAAAAAAAAA  | AAAAA       |             | 635 |
| I K S C    | N .        | K K K K     | K           |             |     |
| L N L      | A T E K    | K K K K     | K           |             |     |
| . I L      | Q L K      | K K K K     | K           |             |     |

5 / 32  
FIG 4

| 10          | 20         | 30          | 40         | 50          |     |
|-------------|------------|-------------|------------|-------------|-----|
| 1234567890  | 1234567890 | 1234567890  | 1234567890 | 1234567890  |     |
| ATGCCCCCTCC | CTTATCATA  | CTTTCTCTTT  | ACTGTCTCT  | TACCCCCCTT  | 50  |
| M A L P     | Y H T      | F L F       | T V L L    | P P F       |     |
| W P S       | L I I L    | F S L       | L F S      | Y P L S     |     |
| G P P       | L S Y      | F S L Y     | C S L      | T P F       |     |
| CGCTCTCACT  | GCACCCGCTC | CATGCTGCTG  | TACAACCAGT | AGCTCCCCCTT | 100 |
| A L T       | A P P P    | C C C       | T T S      | S S P Y     |     |
| L S L       | H P L      | H A A V     | Q P V      | A P L       |     |
| R S H C     | T P S      | M L L       | Y N Q      | L P L       |     |
| ACCAAGAGTT  | TCTATGAAGA | ACCGCGGCTTC | CTGGAATAT  | TGATGCCCCA  | 150 |
| Q E F       | L . R      | T R L P     | G N I      | D A P       |     |
| T K S F     | Y E E      | R G F       | L E I L    | M P H       |     |
| P R V       | S M K N    | A A S       | W K Y      | C P I       |     |
| TCATATAGGA  | GTTTATCTAA | GGGAAACTCC  | ACCTTCACTG | CCCACACCCA  | 200 |
| S Y R S     | L S K      | G N S       | T F T A    | H T H       |     |
| H I G       | V Y L R    | E T P       | P S L      | P T P I     |     |
| I . E       | F I .      | G K L H     | L H C      | P H P       |     |
| TATGCCCCGC  | AACIGCTATA | ACTCTGCCAC  | TCTTTGCATG | CATGCAAATA  | 250 |
| M P R       | N C Y N    | S A T       | L C M      | H A N T     |     |
| C P A       | T A I      | T L P L     | F A C      | M Q I       |     |
| Y A P Q     | L L .      | L C H       | S L H A    | C K Y       |     |
| CTCATTATTG  | GACAGGGAAA | ATGATTAAATC | CTAGTGTGCC | TGGAGGACTT  | 300 |
| H Y W       | T G K      | M I N P     | S C P      | G G L       |     |
| L I I G     | Q G K      | . L I       | L V V L    | E D L       |     |
| S L L       | D R E N    | D . S       | . L S      | W R T W     |     |
| GGAGCCACTG  | TCTGTGGAC  | TTACTTCACC  | CATACCAGTA | TGCTCTGATGG | 350 |
| G A T V     | C W T      | Y F T       | H T S M    | S D G       |     |
| E P L       | S V G L    | T S P       | I P V      | C L M G     |     |
| S H C       | L L D      | L L H P     | Y Q Y      | V . W       |     |

6 / 32  
FIG 4 (continued)

| 10         | 20          | 30         | 40         | 50          |      |
|------------|-------------|------------|------------|-------------|------|
| 1234567890 | 1234567890  | 1234567890 | 1234567890 | 1234567890  |      |
| ACCTCACTG  | TGTAAAATTT  | AGCAATACTA | TAGACACAAC | CAGCTCCCAA  | 750  |
| L T C      | V K F       | S N T I    | D T T      | S S Q       |      |
| T S P V    | . N L       | A I L      | . T Q P    | A P N       |      |
| P H L      | C K I .     | Q Y Y      | R H N      | Q L P M     |      |
| TGCATCAGGT | GGGTAAACACC | TCCCACACGA | ATAGTCTGCC | TACCCCTCAGG | 800  |
| C I R W    | V T P       | P T R      | I V C L    | P S G       |      |
| A S G      | G . H L     | P H E      | . S A      | Y P Q E     |      |
| H Q V      | G N T       | S H T N    | S L P      | T L R       |      |
| AATATTTTTT | GTCGTGGGTA  | CCTCAGCCTA | TCATGTGTTG | AATGGCTCTT  | 850  |
| I F F      | V C G T     | S A Y      | H C L      | N G S S     |      |
| Y F L      | S V V       | P Q P I    | I V .      | M A L       |      |
| N I F C    | L W Y       | L S L      | S L F E    | W L F       |      |
| CAGAATCTAT | GTCCTTCCTC  | TCATTCTTAG | TGCCCCCTAT | GACCATCTAC  | 900  |
| E S M      | C F L       | S F L V    | P P M      | T I Y       |      |
| Q N L C    | A S S       | H S .      | C P L .    | P S T       |      |
| R I Y      | V L P L     | I L S      | A P Y      | D H L H     |      |
| ACTGAACAAG | ATTTATACAA  | TCATGTGGTA | OCTAAGCCCC | ACAACAAAAG  | 950  |
| T E Q D    | L Y N       | H V V      | P K P H    | N K R       |      |
| L N K      | I Y T I     | M S Y      | L S P      | T T K E     |      |
| . T R      | F I Q       | S C R T    | . A P      | Q Q K       |      |
| AGTACCCATT | CTTCCTTTTG  | TTATCAGAGC | AGGAGTGCTA | GGCAGACTAG  | 1000 |
| V P I      | L P F V     | I R A      | G V L      | G R L G     |      |
| Y P F      | F L L       | L S E Q    | E C .      | A D .       |      |
| S T H S    | S F C       | Y Q S      | R S A R    | Q T R       |      |
| GTACTGGCAT | TGGCAGTATC  | ACAACCTCTA | CTCAGTTCTA | CTACAAACTA  | 1050 |
| T G I      | G S I       | T T S T    | Q F Y      | Y K L       |      |
| V L A L    | A V S       | Q P L      | L S S T    | T N Y       |      |
| Y W H      | W Q Y H     | N L Y      | S V L      | L Q T I     |      |

FIG 4 (continued)

| 10         | 20         | 30         | 40          | 50          |      |
|------------|------------|------------|-------------|-------------|------|
| 1234567890 | 1234567890 | 1234567890 | 1234567890  | 1234567890  |      |
| TCTCAAGAAA | TAAATGGTGA | CATGGAACAG | GTCACITGACT | COCTGGTCAC  | 1100 |
| S Q E I    | N G D M    | E Q V T    | D S L V     | T           |      |
| L K K .    | M V T W    | N R S L    | T P W S     | P           |      |
| S R N K    | W . H      | G T G H    | . L P       | G H         |      |
| CTTGCAAGAT | CAACTTAACT | COCTAGCAGC | AGTAGTCCTT  | CAAAATCGAA  | 1150 |
| L Q D Q    | L N S L    | A A V V    | L Q N R     | R           |      |
| C K I N    | L T P .    | Q Q . S    | F K I E     |             |      |
| L A R S    | T . L      | P S S S    | S S P S     | K S K       |      |
| GAGCTTTAGA | CTTGCTAACC | GCCAAAAGAG | GGGGAACCTG  | TTTATTTTTA  | 1200 |
| A L D L    | L T A K    | R G G T    | C L F L     |             |      |
| E L . T    | C . P P    | K E G .    | E P V Y     | F .         |      |
| S F R L    | A N R Q    | K R G N    | L F I F     | R           |      |
| GGAGAAGAAC | GCTGTTATTA | TGTTAATCAA | TCCAGAATTG  | TCACTGAGAA  | 1250 |
| G E E R    | C Y Y V    | N Q S R    | I V T E     | K           |      |
| E K N A    | V I M L    | I N P E    | L S L R     | K           |      |
| R R T L    | L L L C    | . S I Q    | N C H .     | E           |      |
| AGTTAAAGAA | ATTGAGATC  | GAATACAATG | TAGAGCAGAG  | GAGCTTCAAA  | 1300 |
| V K E I    | R D R I    | Q C R A    | E E L Q     | N           |      |
| L K K F    | E I E Y    | N V E Q    | R S F K     |             |      |
| S . R N    | S R S N    | T M . S    | R G A S     | K           |      |
| ACACCGAACG | CTGGGGCCTC | CTCAGCCAAT | GGATGCCCTG  | GGTTCCTCCCC | 1350 |
| T E R W    | G L L S    | Q W M P    | W V L P     |             |      |
| T P N A    | G A S S    | A N G C    | P G F S     | P           |      |
| H R T L    | G P P Q    | P M D A    | L G S P     | L           |      |
| TTCTTAGGAC | CTCTAGCAGC | TCTAATATTG | TTACTCCTCT  | TTGGACCTTG  | 1400 |
| F L G P    | L A A L    | I L L L    | L L F G     | P C         |      |
| S . D L    | . Q L .    | Y C Y S    | S L D P     | V           |      |
| L R T S    | S S S S    | N I V T    | P L W T     | L           |      |

## FIG 4 (continued)

| 10                                                     | 20         | 30         | 40         | 50         |      |   |   |   |   |   |   |   |   |   |   |      |  |
|--------------------------------------------------------|------------|------------|------------|------------|------|---|---|---|---|---|---|---|---|---|---|------|--|
| 1234567890                                             | 1234567890 | 1234567890 | 1234567890 | 1234567890 |      |   |   |   |   |   |   |   |   |   |   |      |  |
| TATCTTTAAC CTCCTTGTTA AGTTTGICTC TTCCAGAATT GAAGCTGTAA |            |            |            |            | 1450 |   |   |   |   |   |   |   |   |   |   |      |  |
| I                                                      | F          | N          | L          | L          | V    | K | F | V | S | S | R | I | E | A | V | K    |  |
| S                                                      | L          | T          | S          | L          | L    | S | L | S | L | P | E | L | K | L | . |      |  |
| Y                                                      | L          | .          | P          | P          | C    | . | V | C | L | F | Q | N | . | S | C | K    |  |
| AGCTACAGAT GGTCTTACAA ATGGAACCCC A                     |            |            |            |            |      |   |   |   |   |   |   |   |   |   |   | 1481 |  |
| L                                                      | Q          | M          | V          | L          | Q    | M | E | P |   |   |   |   |   |   |   |      |  |
| S                                                      | Y          | R          | W          | S          | Y    | K | W | N | P |   |   |   |   |   |   |      |  |
| A                                                      | T          | D          | G          | L          | T    | N | G | T | P |   |   |   |   |   |   |      |  |

| 10                                                     | 20         | 30         | 40         | 50         |     |
|--------------------------------------------------------|------------|------------|------------|------------|-----|
| 1234567890                                             | 1234567890 | 1234567890 | 1234567890 | 1234567890 |     |
| TCAAAATCGA AGAGCTTTAG ACTTGCTAAC CGCCAAAAGA GGGGGAACCT |            |            |            |            | 50  |
| S K S K S F R L A N R Q K R G N L                      |            |            |            |            |     |
| Q N R R A L D L L T A K R G G T C                      |            |            |            |            |     |
| K I E E L . T C . P P K E G E P                        |            |            |            |            |     |
| GTTTATTTTT AGGGGAAGAA TGCTGTTAGT ATGTTAATCA ATCTGGAATC |            |            |            |            | 100 |
| F I F R G R M L L V C . S I W N H                      |            |            |            |            |     |
| L F L G E E C C . Y V N Q S G I                        |            |            |            |            |     |
| V Y F . G K N A V S M L I N L E S                      |            |            |            |            |     |
| ATTACTGAGA AAGTTAAAGA AATTGAGAT CGAATATAAT GTAGAGCAGA  |            |            |            |            | 150 |
| Y . E S . R N L R S N I M . S R                        |            |            |            |            |     |
| I T E K V K E I . D R I . C R A E                      |            |            |            |            |     |
| L L R K L K K F E I E Y N V E Q R                      |            |            |            |            |     |
| GGACCTTCAA AACACTGCAC CCTGGGGGCT CCTCAGCCAA TGGATGCCCT |            |            |            |            | 200 |
| G P S K H C T L G P P Q P M D A L                      |            |            |            |            |     |
| D L Q N T A P W G L L S Q W M P W                      |            |            |            |            |     |
| T F K T L H P G A S S A N G C P                        |            |            |            |            |     |
| GGACTCTCCC CTTCCTAGGA CCTCTAGCAG CTATAATATT TTTACTCTC  |            |            |            |            | 250 |
| D S P L L R T S S S Y N I F T P L                      |            |            |            |            |     |
| T L P F L G P L A A I I F L L L                        |            |            |            |            |     |
| G L S P S . D L . Q L . Y F Y S S                      |            |            |            |            |     |
| TTTGGACCCCT GTATCTTCAA CTTCCTTGIT AAGTTTGICT CTTCAGAAT |            |            |            |            | 300 |
| W T L Y L Q L P C . V C L F Q N                        |            |            |            |            |     |
| F G P C I F N F L V K F V S S R I                      |            |            |            |            |     |
| L D P V S S T S L L S L S L P E L                      |            |            |            |            |     |
| TGAAGCTGTA AAGCTACAAA TAGTTCTTCA AATGGAACCC CAGATGCAGT |            |            |            |            | 350 |
| . S C K A T N S S S N G T P D A V                      |            |            |            |            |     |
| E A V K L Q I V L Q M E P Q M Q S                      |            |            |            |            |     |
| K L . S Y K . F F K W N P R C S                        |            |            |            |            |     |



10 / 32

## FIG 5 (continued)

| 10         | 20         | 30         | 40         | 50         |     |
|------------|------------|------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |     |
| CCATGACTAA | AATCTACCGT | GGACCCCTGG | ACCGGCCTGC | TAGACTATGC | 400 |
| H D .      | N L P W    | T P G      | P A C      | . T M L    |     |
| M T K      | I Y R      | G P L D    | R P A      | R L C      |     |
| P . L K    | S T V      | D P W      | T G L L    | D Y A      |     |
| TCIGATGTTA | ATGACATTGA | AGTCACCCCT | CCCGAGGAAA | TCTCAACTGC | 450 |
| . C .      | . H .      | S H P S    | R G N      | L N C      |     |
| S D V N    | D I E      | V T P      | P E E I    | S T A      |     |
| L M L      | M T L K    | S P L      | P R K      | S Q L H    |     |
| ACAACCCCTA | CTACACTCCA | ATTCAGTAGG | AAGCAGTTAG | ACCAGTTGTC | 500 |
| T T P T    | T L Q      | F S R      | K Q L E    | Q L S      |     |
| Q P L      | L H S N    | S V G      | S S .      | S S C Q    |     |
| N P Y      | Y T P      | I Q .      | E A V R    | A V V      |     |
| AGCCAACCTC | CCCAACAGTA | CTTGGGTTTT | CCTGTTGAGA | GGGTGGACTG | 550 |
| A N L      | P N S T    | W V F      | L L R      | G W T E    |     |
| P T S      | P T V      | L G F S    | C . E      | G G L      |     |
| S Q P P    | Q Q Y      | L G F      | P V E R    | V D .      |     |
| AGAGACAGGA | CTAGCTGGAT | TTCTTAGGCT | GACTAAGAAT | CCCAAGCCT  | 600 |
| R Q D      | . L D      | F L G .    | L R I      | P K P      |     |
| R D R T    | S W I      | S . A      | D . E      | S X S L    |     |
| E T G      | L A G F    | P R L      | T K N      | P X A X    |     |
| ANCTGGGAAG | GTGACCGCAT | CCATCTTTAA | ACATGGGGCT | TGCAACTTAG | 650 |
| X W E G    | D R I      | H L .      | T W G L    | Q L S      |     |
| X G K      | V T A S    | I F K      | H G A      | C N L A    |     |
| L G R      | . P H      | P S L N    | M G L      | A T .      |     |
| CTCACACCCG | ACCAATCAGA | GAGCTCACTA | AAATGCTAAT | CAGGCAAAAA | 700 |
| S H P      | T N Q R    | A H .      | N A N      | Q A K T    |     |
| H T R      | P I R      | E L T K    | M L I      | R Q K      |     |
| L T P D    | Q S E      | S S L      | K C .      | S G K N    |     |

## FIG 5 (continued)

| 10         | 20         | 30         | 40         | 50         |      |
|------------|------------|------------|------------|------------|------|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |      |
| CAGGAGGTAA | AGCAATAGCC | AATCATCTAT | TGCTGAGAG  | CACAGCGGA  | 750  |
| G G K      | A I A      | N H L L    | P E S      | T A G      |      |
| Q E V K    | Q . P      | I I Y      | C L R A    | Q R E      |      |
| R R .      | S N S Q    | S S I A    | . E H S    | G K        |      |
| AGGACAAGGA | TTGGGATATA | AATCAGGCA  | TTCAAGCCAG | CAACAGCAAC | 800  |
| R T R I    | G I .      | T Q A      | F K P A    | T A T      |      |
| G Q G      | L G Y K    | L R H      | S S Q      | Q Q Q P    |      |
| D K D      | W D I      | N S G I    | Q A S      | N S N      |      |
| CCCCTTGGG  | TCCCCCCCCA | TTGTATGGGA | GCTCTGTTTT | CACTCTATTT | 850  |
| P F G      | S P P I    | V W E      | L C F      | H S I S    |      |
| P L G      | P L P      | L Y G S    | S V F      | T L F      |      |
| P L W V    | P S H      | C M G      | A L F S    | L Y F      |      |
| CACTCTATTA | AATCATGCAA | CTGCACTCTT | CTGGTCGGTG | TTTTTATGG  | 900  |
| L Y .      | I M Q      | L H S S    | G P C      | F L W      |      |
| H S I K    | S C N      | C T L      | L V R V    | F Y G      |      |
| T L L      | N H A T    | A L F      | W S V      | F F M A    |      |
| CTCAAGCTGA | GCTTTTGTTC | GCCATCCACC | ACTGCTGTTT | GCCACGGTCA | 950  |
| L K L S    | F C S      | P S T      | T A V C    | H R H      |      |
| S S .      | A F V R    | H P P      | L L F      | A T V T    |      |
| Q A E      | L L F      | A I H H    | C C L      | P P S      |      |
| CAGACCGCT  | GCTGACTTCC | ATCCCTTTGG | ATCCAGCAGA | GTTGCCACTG | 1000 |
| R P A      | A D F H    | P F G      | S S R      | V S T V    |      |
| D P L      | L T S      | I P L D    | P A E      | C P L      |      |
| Q T R C    | . L P      | S L W      | I Q Q S    | V H C      |      |
| TGCTCTGAT  | CCAGCGAGGT | ACCCATTGCC | ACTCCCGATC | AGGCTAAAGG | 1050 |
| L L I      | Q R G      | T H C H    | S R S      | G . R      |      |
| C S .      | S S E V    | P I A      | T P D Q    | A K G      |      |
| A P D      | P A R Y    | P L P      | L P I      | R L K A    |      |

12 / 32

## FIG 5 (continued)

| 10         | 20         | 30         | 40          | 50         |      |
|------------|------------|------------|-------------|------------|------|
| 1234567890 | 1234567890 | 1234567890 | 1234567890  | 1234567890 |      |
| CTTGCCATTG | TTCTGCGATG | GCTAAGTGGC | TGGGTTTGTG  | CTAATAGAAC | 1100 |
| L A I V    | P A W      | L S A      | W V C P     | N R T      |      |
| L P L      | F L H G    | . V P      | G F V       | L I E L    |      |
| C H C      | S C M      | A K C L    | G L S       | . . N      |      |
|            |            |            |             |            |      |
| TGAACACTGG | TCACTGGGTT | CCATGGTTCT | CTTCCATGAC  | CCACGGCTTC | 1150 |
| E H W      | S L G S    | M V L      | F H D       | P R L L    |      |
| N T G      | H W V      | P W F S    | S M T       | H G F      |      |
| . T L V    | T G F      | H G S      | L P .       | P T A S    |      |
|            |            |            |             |            |      |
| TAATAGAGCT | ATAACACTCA | CCGCATGGCC | CAAGATTCCA  | TTCTTTGGTA | 1200 |
| I E L      | . H S      | P H G P    | R F H       | S L V      |      |
| . . S Y    | N T H      | R M A      | Q D S I     | P W Y      |      |
| N R A      | I T L T    | A W P      | K I P       | F L G I    |      |
|            |            |            |             |            |      |
| TCTGTGAGGC | CAAGAAACCC | AGGTCAGAGA | ANGTGAGGCT  | TGCCACCATT | 1250 |
| S V R P    | R T P      | G Q R      | X . G L     | P P F      |      |
| L . G      | Q E P Q    | V R E      | X E A       | C H H L    |      |
| C E A      | K N P      | R S E X    | V R L       | A T I      |      |
|            |            |            |             |            |      |
| TGGGAAGTGG | CCCACTGCCA | TTTGGTACG  | GGCCCCACCAC | CATCTTGGGA | 1300 |
| G K W      | P T A I    | L V A      | A H H       | H L G S    |      |
| G S G      | P L P      | F W .      | R P T T     | I L G      |      |
| W E V A    | H C H      | F G S      | G P P P     | S W E      |      |
|            |            |            |             |            |      |
| GCTGTGGGAG | CAAGGATCCC | CCAGTAACA  |             |            | 1329 |
| C G S      | K D P      | P V T      |             |            |      |
| A V G A    | R I P      | Q .        |             |            |      |
| L W E      | Q G S P    | S N        |             |            |      |

13 / 32

FIG 6

| 10         | 20         | 30          | 40          | 50         |     |
|------------|------------|-------------|-------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890  | 1234567890  | 1234567890 |     |
| CCTAGAACGT | ATTCTGGAGA | ATTGGGACCA  | ATGTGACACT  | CAGACCTAA  | 50  |
| P R T Y    | S G E L    | G P M .     | H S .       | D A K      |     |
| L E R      | I L E N    | W D Q       | C D T       | Q T L R    |     |
| . N V      | F W R      | I G T N     | V T L       | R R .      |     |
| GAAAGAAAG  | ATTATATTC  | TTCCTGCAGTA | CGGCTGGGC   | ACAATATCCT | 100 |
| K E T      | I Y I L    | L Q Y       | R L A       | T I S S    |     |
| K K R      | F I F      | F C S T     | A W P       | Q Y P      |     |
| E R N D    | L Y S      | S A V       | P P G H     | N I L      |     |
| CTTCAAGGA  | GAGAAACCTG | GCTTCTGAG   | GGAAGTATAA  | ATTATAACAT | 150 |
| S R E      | R N L A S  | . G         | K Y K L .   | H          |     |
| L Q G R    | E T W      | L P E       | G S I N     | Y N I      |     |
| F K G      | E K P G    | F L R       | E V .       | I I T S    |     |
| CATCTACAG  | CTAGACCTCT | TCGTAGAAA   | GGAGGGCAA   | TGGAGTGAAG | 200 |
| H L T A    | R P L L .  | K G G       | Q M E .     | S          |     |
| I L Q      | L D L F    | C R K       | E G K       | W S E V    |     |
| S Y S .    | T S        | S V E R     | R A N       | G V K      |     |
| TGCCATATGT | GCAAACTTTC | TTTTCATTA   | GAGACAACCTC | ACAATTATGT | 250 |
| A I C      | A N F L    | F I K       | R Q L       | T I M .    |     |
| P Y V      | Q T F      | F S L R     | D N S       | Q L C      |     |
| C H M C    | K L S      | F H .       | E T T H     | N Y V      |     |
| AAAAAGTGTG | GTTTATGCCC | TACAGGAAGC  | CCTCAGAGTC  | CACCTGCTTA | 300 |
| K V W      | F M P      | Y R K P     | S E S       | T S L      |     |
| K K C G    | L C P      | T G S       | P Q S P     | P P Y      |     |
| K S V      | V Y A L    | Q E A       | L R V       | H L P T    |     |
| CCCCAGGTC  | CCCTCCCCGA | CTCCTTCTC   | AACTAATAAG  | GACCCCCCTT | 350 |
| P Q R P    | L P D      | S F L N . . | G P P F     |            |     |
| P S V      | P S P T    | P S S       | T N K       | D P P L    |     |
| P A S      | P P R      | L L P Q     | L I R       | T P L      |     |
| TAAACCAAC  | GGTCCAAAG  | GAGATAGACA  | AAGGGGTAAA  | CAATGAACCA | 400 |
| N P N      | G P K G    | D R Q       | R G K Q .   | T K        |     |
| T Q T      | V Q K      | E I D K     | G V N       | N E P      |     |
| . P K R    | S K R R .  | T K G .     | T M N Q     |            |     |
| AAGAGTGCCA | ATATTCCCCG | ATTATGCCCC  | CTCCAAGCAG  | TGAGAGGAGG | 450 |
| E C Q      | Y S P      | I M P P     | P S S       | E R R      |     |
| K S A N    | I P R      | L C P       | L Q A V     | R G G      |     |
| R V P      | I F P D    | Y A P       | S K Q .     | E E E      |     |
| AGAATTGGC  | CCAGCCAGAG | TGCTGTACC   | TTTTTCTCTC  | TCAGACTTAA | 500 |
| R I R P    | S Q S      | A C T       | F F S L     | R L K      |     |
| E F G      | P A R V    | P V P       | F S L       | S D L K    |     |
| N S A      | Q P E      | C L Y L     | F L S       | Q T .      |     |

14 / 32

## FIG 6 (continued)

| 10         | 20         | 30          | 40          | 50            |      |
|------------|------------|-------------|-------------|---------------|------|
| 1234567890 | 1234567890 | 1234567890  | 1234567890  | 1234567890    |      |
| AGCAAATTAA | AATAGACCTA | GGTAAATTCT  | CAGATAACCC  | TGACGGCTAT    | 550  |
| A N .      | N R P R .  | I L R .     | P .         | R L Y         |      |
| Q I K      | I D L      | G K F S     | D N P       | D G Y         |      |
| S K L K .  | T .        | V N S       | Q I T L     | T A I         |      |
| ATTGATGTTT | TACAAGGGTT | AGGACAATCC  | TTTGATCTGA  | CATGGAGAGA    | 600  |
| . C F      | T R V      | R T I L     | . S D       | M E R         |      |
| I D V L    | Q G L      | G Q S       | F D L T     | W R D         |      |
| L M F      | Y K G .    | D N P       | L I .       | H G E I       |      |
| TATAATGTTA | CTACTAAATC | AGACACTAAC  | CCCAATGAG   | AGAAGTGGCG    | 650  |
| Y N V T    | T K S      | D T N       | P K .       | E K C R       |      |
| I M L      | L L N Q    | T L T       | P N E       | R S A A       |      |
| . C Y Y .  | I R H .    | P           | Q M R       | E V P         |      |
| CIGTAACTGC | AGCCCGAGAG | TTTGGCGATC  | TTTGGTATCT  | CAGTCAGGCC    | 700  |
| C N C      | S P R V    | W R S       | L V S       | Q S G Q       |      |
| V T A      | A R E      | F G D L     | W Y L       | S Q A         |      |
| L .        | L Q        | P E S       | L A I       | F G I S V R P |      |
| AACAATAGGA | TGACAACAGA | GGAAAGAACA  | ACTCCCAACAG | GCCAGCAGGC    | 750  |
| Q .        | D D N R    | G K N N     | S H R       | P A G         |      |
| N N R M    | T T E      | E R T       | T P T G     | Q Q A         |      |
| T I G .    | Q Q R      | K E Q       | L P Q       | A S R Q       |      |
| AGTTCCCACT | GTAGACCTTC | ATTGGGACAC  | AGAATCAGAA  | CATGGAGATT    | 800  |
| S S Q C    | R P S      | L G H       | R I R T     | W R L         |      |
| V P S      | V D P H    | W D T       | E S E       | H G D W       |      |
| F P V .    | T L        | I G T Q     | N Q N       | M E I         |      |
| GGTGGCACA  | ACATTGCTA  | ACTTGGGTGC  | TAGAAGGACT  | GAGGAAAAT     | 850  |
| V P Q      | T F A N    | L R A       | R R T       | E E N .       |      |
| C H K      | H L L      | T C V L     | E G L       | R K T         |      |
| G A T N    | I C .      | L A C .     | K D .       | G K L         |      |
| AGGAGAACG  | CTATGAATTA | CITCAATGATG | TCCACTATAA  | CACAGGGAA     | 900  |
| E E A      | Y E L      | L N D V     | H Y N       | T G K         |      |
| R K K P    | M N Y      | S M M       | S T I T     | Q G K         |      |
| G R S      | L .        | I T Q .     | C P L .     | H R E R       |      |
| GGAAGAAAT  | CTTACTGCTT | TTCTGGACAG  | ACTAAGGGAG  | GCATTGAGGA    | 950  |
| G R K S    | Y C F      | S G Q       | T K G G     | I E E         |      |
| E E N      | L T A F    | L D R       | L R E       | A L R K       |      |
| K K I      | L L L      | F W T D .   | G R H .     | G             |      |
| AGCATACCTC | CCGTCACT   | GACTCTATTG  | AAGGCCAACT  | AATCTTAAAG    | 1000 |
| A Y L      | P V T .    | L Y .       | R P T       | N L K G       |      |
| H T S      | L S P      | D S I E     | G Q L       | I L K         |      |
| S I P P    | C H L      | T L L       | K A N .     | S . R         |      |

15 / 32  
FIG 6 (continued)

| 10         | 20         | 30         | 40         | 50         |      |
|------------|------------|------------|------------|------------|------|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |      |
| GATAAGTTTA | TCACTCAGTC | AGCTGCAGAC | ATTAGAAAAA | ACTTCAAAG  | 1050 |
| . V Y      | H S V      | S C R H    | . K K      | L Q K      |      |
| D K F I    | T Q S      | A A D      | I R K N    | F K S      |      |
| I S L      | S L S Q    | L Q T      | L E K      | T S K V    |      |
| TCTGCTTAG  | GGCGGAGCA  | GAACTTAGAA | ACCTTATTTA | ACTTGGCATC | 1100 |
| S A L G    | P E Q      | N L E      | T L F N    | L A S      |      |
| L P .      | A R S R    | T . K      | P Y L      | T W H P    |      |
| C L R      | P G A      | E L R N    | P I .      | L G I      |      |
| CTCAGTTTTT | TATAATAGAG | ATCAGGAGGA | GCAGGCGAAA | CGGACAAAC  | 1150 |
| S V F      | Y N R D    | Q E E      | Q A K      | R D K R    |      |
| Q F F      | I I E      | I R R S    | R R N      | G T N      |      |
| L S F L    | . . R      | S G G      | A G E T    | G Q T      |      |
| GGGATAAAAA | AAAAAGGGGG | GGTCCACTAC | TTTAGTCATG | GGCTCAGGC  | 1200 |
| D K K      | K R G      | G P L L    | . S W      | P S G      |      |
| G I K K    | K G G      | V H Y      | F S H G    | P Q A      |      |
| G . K      | K K G G    | S T T      | L V M      | A L R Q    |      |
| AAGCAGACTT | TGGAGGCTCT | GCAAAAGGGA | AAAGCTGGGC | AAATCAAATG | 1250 |
| K Q T L    | E A L      | Q K G      | K A G Q    | I K C      |      |
| S R L      | W R L C    | K R E      | K L G      | K S N A    |      |
| A D F      | G G S      | A K G K    | S W A      | N Q M      |      |
| OCTAATAGGG | CTGGCTTCCA | GTGGGGTCTA | CAAGGACACT | TTAAAAAGA  | 1300 |
| L I G      | L A S S    | A V Y      | K D T      | L K K I    |      |
| . . G      | W L P      | V R S T    | R T L      | . K R      |      |
| P N R A    | G F Q      | C G L      | Q G H F    | K K D      |      |
| TTATCCAAGT | AGAAATAAGC | GGGGGGCTTG | TCCATGCCCC | TTACGTCAAG | 1350 |
| I Q V      | E I S      | R P L V    | H A P      | Y V K      |      |
| L S K .    | K . A      | A P L      | S M P L    | T S R      |      |
| Y P S      | R N K P    | P P C      | P C P      | L R Q G    |      |
| GGATCACTG  | GAAGGCCCAC | TGCCCCAGGG | GATGAAGATA | CTCTGAGTCA | 1400 |
| G I T G    | R P T      | A P G      | D E D T    | L S Q      |      |
| E S L      | E G P L    | P Q G      | M K I      | L . V R    |      |
| N H W      | K A H      | C P R G    | . R Y      | S E S      |      |
| GAAGCCATTA | ACCAGATGAT | CCAGCAGCAG | GACTGAGGGT | GGGGGGGGG  | 1450 |
| K P L      | T R .      | S S S R    | T E G      | A R G E    |      |
| S H .      | P D D      | P A A G    | L R V      | P G A      |      |
| E A I N    | Q M I      | Q Q Q      | D . G C    | P G R      |      |
| AGGGGAGGC  | CATGCCATCA | CCCTCAGAGA | GGGGGGGGTA | TGTTTGACCA | 1500 |
| R Q P      | M P S      | P S Q S    | P G Y      | V . P      |      |
| S A S P    | C H H      | P H R      | A P G M    | F D H      |      |
| A P A      | H A I T    | L T E      | P R V      | C L T I    |      |

## FIG 6 (continued)

| 10           | 20         | 30         | 40         | 50         |
|--------------|------------|------------|------------|------------|
| 1234567890   | 1234567890 | 1234567890 | 1234567890 | 1234567890 |
| TTGAGAGCCA A |            |            |            |            |
| L R A        |            |            |            | 1511       |
| . E P        |            |            |            |            |
| E S Q        |            |            |            |            |

FIG 7

| 10         | 20         | 30          | 40         | 50         |     |
|------------|------------|-------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890  | 1234567890 | 1234567890 |     |
| ATGGGCAGCA | GCCATCATCA | TCATCATCAC  | AGCAGGGGCC | TGGTGCCGCG | 50  |
| M G S S    | H H H      | H H H       | S S G L    | V P R      |     |
| CGGCAGCCAT | ATGGCTAGCA | TGACTGGTGG  | ACAGCAAATG | GGTCGGATCC | 100 |
| G S H      | M A S M    | T G G       | Q Q M      | G R I L    |     |
| TAGAACTAT  | TCTGGAGAAT | TGGGACCAAT  | GTGACACTCA | GACGCTAAGA | 150 |
| E R I      | L E N      | W D Q C     | D T Q      | T L R      |     |
| AAGAAAGAT  | TTATATTCTT | CTGCAGTACC  | GCCTGGCCAC | AATATCCTCT | 200 |
| K K R F    | I F F      | C S T       | A W P Q    | Y P L      |     |
| TCAAGGGAGA | GAAACCTGGC | TTCTTGAGGG  | AAGTATAAAT | TATAACATCA | 250 |
| Q G R      | E T W L    | P E G       | S I N      | Y N I I    |     |
| TCTTACAGCT | AGACCTCTTC | TGTAGAAAGG  | AGGGCAAATG | GAGTGAAGTG | 300 |
| L Q L      | D L F      | C R K E     | G K W      | S E V      |     |
| CCATATGTGC | AAACTTTCTT | TTCAATTAAGA | GACAACTCAC | AATTATGTAA | 350 |
| P Y V Q    | T F F      | S L R       | D N S Q    | L C K      |     |
| AAAGTGTGGT | TTATGCCCTA | CAGGAAGCCC  | TCAGAGTCCA | CCTCCCTACC | 400 |
| K C G      | L C P T    | G S P       | Q S P      | P P Y P    |     |
| CCAGGGTCCC | CTCCCCGACT | CCTTCTCTCA  | CTAATAAGGA | CCCCCCTTTA | 450 |
| S V P      | S P T      | P S S T     | N K D      | P P L      |     |
| ACCCAAACGG | TCCAAAAGGA | GATAGACAAA  | GGGGTAAACA | ATGAACCAAA | 500 |
| T Q T V    | Q K E      | I D K       | G V N N    | E P K      |     |
| GAGTGCCAT  | ATTCCCCGAT | TATGCCCCCT  | CCAAGCAGTG | AGAGGAGGAG | 550 |
| S A N      | I P R L    | C P L       | Q A V      | R G G E    |     |
| AATTGGGCCC | AGCCAGAGTG | CTGTACCTT   | TTTCTCTCTC | AGACTTAAG  | 600 |
| F G P      | A R V      | P V P F     | S L S      | D L K      |     |
| CAAATTAAAA | TAGACCTAGG | TAAATTCTCA  | GATAACCTTG | ACGGCTATAT | 650 |
| Q I K I    | D L G      | K F S       | D N P D    | G Y I      |     |
| TGATGTTTIA | CAAGGGTTAG | GACAATCCTT  | TGATCTGACA | TGGAGAGATA | 700 |
| D V L      | Q G L G    | Q S F       | D L T      | W R D I    |     |
| TAATGTACT  | ACTAAATCAG | ACACTAACCC  | CAAATGAGAG | AAGTGCCGCT | 750 |
| M L L      | L N Q      | T L T P     | N E R      | S A A      |     |
| GTAAGTGCAG | CCCGAGAGTT | TGGCGATCTT  | TGGTATCTCA | GTGAGGOCAA | 800 |
| V T A A    | R E F      | G D L       | W Y L S    | Q A N      |     |



## FIG 7 (continued)

| 10                                                      | 20         | 30         | 40         | 50         |      |
|---------------------------------------------------------|------------|------------|------------|------------|------|
| 1234567890                                              | 1234567890 | 1234567890 | 1234567890 | 1234567890 |      |
| CAATAGGATG                                              | ACAACAGAGG | AAAGAACAAC | TCCCACAGGC | CAGCAGGCAG | 850  |
| N R M                                                   | T T E E    | R T T      | P T G      | Q Q A V    |      |
| TCCCCAGTGT AGACCCCTCAT TGGGACACAG AATCAGAACA TGGAGATTGG |            |            |            |            | 900  |
| P S V                                                   | D P H      | W D T E    | S E H      | G D W      |      |
| TGCCACAAAC ATTTCGCTAAC TTGGGTGCTA GAAGGACTGA GGAAACTAG  |            |            |            |            | 950  |
| C H K H                                                 | L L T      | C V L      | E G L R    | K T R      |      |
| GAAGAAGCCT ATGAATTACT CAATGATGTC CACTATAACA CAGGGAAAGG  |            |            |            |            | 1000 |
| K K P                                                   | M N Y S    | M M S      | T I T      | Q G K E    |      |
| AAGAAAATCT TACTGCTTTT CTGGACAGAC TAAGGGAGGC ATTGAGGAAG  |            |            |            |            | 1050 |
| E N L                                                   | T A F      | L D R L    | R E A      | L R K      |      |
| CATACTCCG TGTCACCTGA CTCTATTGAA GGCCAACTAA TCTTAAAGGA   |            |            |            |            | 1100 |
| H T S L                                                 | S P D      | S I E      | G Q L I    | L K D      |      |
| TAAGTTTATC ACTCAGTCAG CTGCAGACAT TAGAAAAAAC TTCAAAAGTC  |            |            |            |            | 1150 |
| K F I                                                   | T Q S A    | A D I      | R K N      | F K S L    |      |
| TGCTTAAGCT TGCGGCGGCA CTGAGCACC ACCACCACCA CCACTGAGAT   |            |            |            |            | 1200 |
| P K L                                                   | A A A      | L E H H    | H H H      | H . D      |      |
| CCGGCTGCTA ACAAAGCCCG AAAGGAAGCT GAGTTGGCTN GTGGCNA     |            |            |            |            | 1247 |
| P A A N                                                 | K A R      | K E A      | E L A X    | G          |      |

19 / 32  
FIG 8

| 10         | 20         | 30         | 40         | 50         |     |
|------------|------------|------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |     |
| ATGGCTAGCA | TGACTGGTGG | ACAGCAAATG | GGTGGGATCC | TAGAACGTAT | 50  |
| M A S M    | T G G      | Q Q M      | G R I L    | E R I      |     |
| TCTGGAGAAT | TGGGACCAAT | GTGACACTCA | GAGGCTAAGA | AAGAAACGAT | 100 |
| L E N      | W D Q C    | D T Q      | T L R      | K K R F    |     |
| TTATATTCTT | CTGCAGTACC | GCCTGGCCAC | AATATCCTCT | TCAAGGGAGA | 150 |
| I F F      | C S T      | A W P Q    | Y P L      | Q G R      |     |
| GAAACCTGGC | TTCCTGAGGG | AAGTATAAAT | TATAACATCA | TCTTACAGCT | 200 |
| E T W L    | P E G      | S I N      | Y N I I    | L Q L      |     |
| AGACCTCTTC | TGTAGAAAGG | AGGGCAAATG | GAGTGAAGTG | CCATATGTGC | 250 |
| D L F      | C R K E    | G K W      | S E V      | P Y V Q    |     |
| AAACTTTCTT | TTCATTAAGA | GACAACTCAC | AATTATGTAA | AAAGTGTGGT | 300 |
| T F F      | S L R      | D N S Q    | L C K      | K C G      |     |
| TTATGCOCTA | CAGGAAGCCC | TCAGAGTCCA | CCTCCCTACC | CCAGGGTCCC | 350 |
| L C P T    | G S P      | Q S P      | P P Y P    | S V P      |     |
| CTCCCGGACT | CCTTCCTCAA | CTAATAAGGA | CCCCCCTTTA | ACCCAAACGG | 400 |
| S P T      | P S S T    | N K D      | P P L      | T Q T V    |     |
| TCCAAAAGGA | GATAGACAAA | GGGGTAAACA | ATGAACCAAA | GAGTGGCAAT | 450 |
| Q K E      | I D K      | G V N N    | E P K      | S A N      |     |
| ATTCCCGGAT | TATGCCCCCT | CCAAGCAGTG | AGAGGAGGAG | AATTGGGCCC | 500 |
| I P R L    | C P L      | Q A V      | R G G E    | F G P      |     |
| AGCCAGAGTG | CCGTACCTTT | TTTCTCTCTC | AGACTTAAAG | CAAATTAAAA | 550 |
| A R V      | P V P F    | S L S      | D L K      | Q I K I    |     |
| TAGACCTAGG | TAAATCTICA | GATAACCTTG | ACGGCTATAT | TGATGTTTTA | 600 |
| D L G      | K F S      | D N P D    | G Y I      | D V L      |     |
| CAAGGGTTAG | GACAATCCTT | TGATCTGACA | TGGAGAGATA | TAATGTTACT | 650 |
| Q G L G    | Q S F      | D L T      | W R D I    | M L L      |     |
| ACTAATCAG  | ACACTAACC  | CAAATGAGAG | AAGTGGCGCT | GTAAGTGCAG | 700 |
| L N Q      | T L T P    | N E R      | S A A      | V T A A    |     |
| CCCGAGAGTT | TGGGATCTTT | TGGTATCTCA | GTGAGGCCAA | CAATAGGATG | 750 |
| R E F      | G D L      | W Y L S    | Q A N      | N R M      |     |
| ACAACAGAGG | AAAGAACAAC | TCCACAGGC  | CAGCAGGCAG | TTCOCAGTGT | 800 |
| T T E E    | R T T      | P T G      | Q Q A V    | P S V      |     |

20 / 32

## FIG 8 (continued)

| 10          | 20          | 30         | 40         | 50         |      |
|-------------|-------------|------------|------------|------------|------|
| 1234567890  | 1234567890  | 1234567890 | 1234567890 | 1234567890 |      |
| AGACCCATCAT | TGGGACACAG  | AATCAGAACA | TGGAGATTGG | TGCCACAAAC | 850  |
| D P H       | W D T E     | S E H      | G D W      | C H K H    |      |
| ATTTCCTAAC  | TTGGGTGCTA  | GAAGGACTGA | GGAAAAGTAG | GAAGAAGCCT | 900  |
| L L T       | C V L       | E G L R    | K T R      | K K P      |      |
| ATGAATTACT  | CAATGATGTC  | CACTATAACA | CAGGGAAAGG | AAGAAAATCT | 950  |
| M N Y S     | M M S       | T I T      | Q G K E    | E N L      |      |
| TACTGCTTTT  | CTGGACAGAC  | TAAGGGAGGC | ATTGAGGAAG | CATACCTCCC | 1000 |
| T A F       | L D R L     | R E A      | L R K      | H T S L    |      |
| TGTCACCTGA  | CCTCTATTGAA | GGCCAACTAA | TCTTAAAGGA | TAAGTTTATC | 1050 |
| S P D       | S I E       | G Q L I    | L K D      | K F I      |      |
| ACTCAGTCAG  | CTGCAGACAT  | TAGAAAAAAC | TTCAAAAGTC | TGCCTAAGCT | 1100 |
| T Q S A     | A D I       | R K N      | F K S L    | P K L      |      |
| TGCGGCGCGA  | CTCGAGCACC  | ACCACCACCA | CCACTGAGAT | CCGGCTGCTA | 1150 |
| A A A       | L E H H     | H H H      | H . D      | P A A N    |      |
| ACAAAGCCCG  | AAAGGAAGCT  | GAGTTGGCTG | GTGGCA     |            | 1186 |
| K A R       | K E A       | E L A      | G G        |            |      |

| 10         | 20         | 30         | 40          | 50         |     |
|------------|------------|------------|-------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890  | 1234567890 |     |
| TGTCGCTGT  | GCTCCTGATC | CAGCACAGGC | GCCCATTTGCC | TCTCCCAATT | 50  |
| C P L C    | S . S      | S T G      | A H C L     | S Q L      |     |
| V R C      | A P D P    | A Q A      | P I A       | S P N W    |     |
| S A V      | L L I      | Q H R R    | P L P       | L P I      |     |
| GGGCTAAAGG | CTTGCCATTG | TTCTTGCACA | GCTAAGTGGC  | TGGGTTCATC | 100 |
| G . R      | L A I V    | P A Q      | L S A       | W V H P    |     |
| A K G      | L P L      | F L H S    | . V P       | G F I      |     |
| G L K A    | C H C      | S C T      | A K C L     | G S S      |     |
| CTAATCGAGC | TGAACACTAG | TCACTGGGTT | CCACGGTTCT  | CTTCCATGAC | 150 |
| N R A      | E H .      | S L G S    | T V L       | F H D      |     |
| L I E L    | N T S      | H W V      | P R F S     | S M T      |     |
| . S S      | . T L V    | T G F      | H G S       | L P . P    |     |
| CCATGGCTTC | TAATAGAGCT | ATAACACTCA | CTGCATGGTC  | CAAGATTCCA | 200 |
| P W L L    | I E L      | . H S      | L H G P     | R F H      |     |
| H G F      | . . S Y    | N T H      | C M V       | Q D S I    |     |
| M A S      | N R A      | I T L T    | A W S       | K I P      |     |
| TTCTTTGGAA | TCCGTGAGAC | CAAGAACCCC | AGGTCAGAGA  | ACACAAGGCT | 250 |
| S L E      | S V R P    | R T P      | G Q R       | T Q G L    |     |
| P W N      | P . D      | Q E P Q    | V R E       | H K A      |     |
| F L G I    | R E T      | K N P      | R S E N     | T R L      |     |
| TGCCACCATG | TTGGAAGCAG | CCCACCACCA | TTTIGGAAGC  | AGCCCGCCAC | 300 |
| P P C      | W K Q      | P T T I    | L E A       | A R H      |     |
| C H H V    | G S S      | P P P      | F W K Q     | P A T      |     |
| A T M      | L E A A    | H H H      | F G S       | S P P L    |     |
| TATCTTGGGA | GCTCTGGGAG | CAAGGACCCC | AGGTAAACAAT | TTGGTGACCA | 350 |
| Y L G S    | S G S      | K D P R    | . Q F       | G D H      |     |
| I L G      | A L G A    | R T P      | G N N       | L V T T    |     |
| S W E      | L W E      | Q G P Q    | V T I       | W . P      |     |
| CGAAGGGACC | TGAATCGCA  | ACCATGAAGG | GATCTOCAA   | GCAATTGGAA | 400 |
| E G T      | . I R N    | H E G      | I S K       | A I G N    |     |
| K G P      | E S A      | T M K G    | S P K       | Q L E      |     |
| R R D L    | N P Q      | P . R      | D L Q S     | N W K      |     |

22 / 32

## FIG 9 (continued)

| 10          | 20         | 30          | 40         | 50         |     |
|-------------|------------|-------------|------------|------------|-----|
| 1234567890  | 1234567890 | 1234567890  | 1234567890 | 1234567890 |     |
| ATGTTCTCTCC | CAAGGCAAAA | ATGCCCCCTAA | GATGTATTCT | GGAGAATTGG | 450 |
| V P P       | K A K      | M P L R     | C I L      | E N W      |     |
| M F L P     | R Q K      | C P .       | D V F W    | R I G      |     |
| C S S       | Q G K N    | A P K       | M Y S      | G E L G    |     |
| GACCAATTTG  | ACCTCAGAC  | AGTAAGAAAA  | AAATGACTTA | TATTCTTCTG | 500 |
| D Q F D     | P Q T      | V R K       | K . L I    | F F C      |     |
| T N L       | T L R Q    | . E K       | N D L      | Y S S A    |     |
| P I .       | P S D      | S K K K     | M T Y      | I L L      |     |
| CAGTACCGCC  | CTGGCCACGA | TATCCTCTTC  | AAGGGGGAGA | AACCTGGCCT | 550 |
| S T A       | L A T I    | S S S       | R G R      | N L A S    |     |
| V P P       | W P R      | Y P L Q     | G G E      | T W P      |     |
| Q Y R P     | G H D      | I L F       | K G E K    | P G L      |     |
| CCTGAGGGAA  | GTATAAATTA | TAACACCATC  | TTACAGCTAG | ACCTGTTTIG | 600 |
| . G K       | Y K L      | . H H L     | T A R      | P V L      |     |
| P E G S     | I N Y      | N T I       | L Q L D    | L F C      |     |
| L R E V     | . I I      | T P S       | Y S .      | T C F V    |     |
| TAGAAAAGGA  | GGCAAATGGA | GTCGAGTGCC  | ATATTTACAA | ACTTTCTTTT | 650 |
| . K R R     | Q M E      | . S A       | I F T N    | F L F      |     |
| R K G       | G K W S    | E V P       | Y L Q      | T F F S    |     |
| E K E       | A N G      | V K C H     | I Y K      | L S F      |     |
| CATTAAAGA   | CAACTCGCAA | TTATGTTAAC  | AGTGTGATTT | GIGTTCTTAC | 700 |
| I K R       | Q L A I    | M L T       | V . F      | V F L H    |     |
| L K D       | N S Q      | L C .       | Q C D L    | C S Y      |     |
| H . K T     | T R N      | Y V N       | S V I C    | V P T      |     |
| ACGGAAGCCC  | TCAGATTCTA | CTCCCCACCC  | CCGGCATCTC | CCCTGAATCC | 750 |
| G S P       | Q I L      | L P T P     | G I S      | P E S      |     |
| T E A L     | R F Y      | S P P       | P A S P    | L N P      |     |
| R K P       | S D S T    | P H P       | R H L      | P . I P    |     |
| CTCCCCAACT  | TATT       |             |            |            | 764 |
| L P N L     |            |             |            |            |     |
| S P T       | Y          |             |            |            |     |
| P Q L       | I          |             |            |            |     |

23 / 32

FIG 10

| 10         | 20         | 30         | 40          | 50          |     |
|------------|------------|------------|-------------|-------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890  | 1234567890  |     |
| TGTCGCTGT  | GCTCCTGATC | CAGCACAGGC | GCCCATGTGCC | TCTCCCAATT  | 50  |
| C P L C    | S . S      | S T G      | A H C L     | S Q L       |     |
| V R C      | A P D P    | A Q A      | P I A       | S P N W     |     |
| S A V      | L L I      | Q H R R    | P L P       | L P I       |     |
| GGGCTAAAGG | CTTGCCATTG | TTCTTGCACA | GCTAAGTGGC  | TGGGTTCATC  | 100 |
| G . R      | L A I V    | P A Q      | L S A       | W V H P     |     |
| A K G      | L P L      | F L H S    | . V P       | G F I       |     |
| G L K A    | C H C      | S C T      | A K C L     | G S S       |     |
| CTAATCGAGC | TGAACACTAG | TCACTGGGTT | CCACGGTTCCT | CTTCCATGAC  | 150 |
| N R A      | E H .      | S L G S    | T V L       | F H D       |     |
| L I E L    | N T S      | H W V      | P R F S     | S M T       |     |
| . S S      | . T L V    | T G F      | H G S       | L P . P     |     |
| CCATGGCTTC | TAATAGAGCT | ATAACACTCA | CTGCAATGGTC | CAAGATTCCA  | 200 |
| P W L L    | I E L      | . H S      | L H G P     | R F H       |     |
| H G F      | . . S Y    | N T H      | C M V       | Q D S I     |     |
| M A S      | N R A      | I T L T    | A W S       | K I P       |     |
| TTCTTTGGAA | TCCGTGAGAC | CAAGAACCCC | AGGTCAGAGA  | ACACAAGGCT  | 250 |
| S L E      | S V R P    | R T P      | G Q R       | T Q G L     |     |
| P W N      | P . D      | Q E P Q    | V R E       | H K A       |     |
| F L G I    | R E T      | K N P      | R S E N     | T R L       |     |
| TGCCACCATG | TTGGAAGCAG | CCCACCAACA | TTTGTGAAGC  | GGCCCCGCCAC | 300 |
| P P C      | W K Q      | P T T I    | L E A       | A R H       |     |
| C H H V    | G S S      | P P P      | F W K R     | P A T       |     |
| A T M      | L E A A    | H H H      | F G S       | G P P L     |     |
| TATCTTGGGA | GCTCTGGGAG | CAAGAACCCC | CAGGTAAACA  | TTTGGTGAAC  | 350 |
| Y L G S    | S G S      | K D P      | Q V T I     | W . P       |     |
| I L G      | A L G A    | R T P      | R . Q       | F G D H     |     |
| S W E      | L W E      | Q G P P    | G N N       | L V T       |     |
| ACGAAGGGAC | CTGAATCCGC | AACCATGAAG | GGATCTCCAA  | AGCAATTGGA  | 400 |
| R R D      | L N P Q    | P . R      | D L Q       | S N W K     |     |
| E G T      | . I R      | N H E G    | I S K       | A I G       |     |
| T K G P    | E S A      | T M K      | G S P K     | Q L E       |     |

24 / 32  
FIG 10 (continued)

| 10         | 20         | 30          | 40         | 50         |     |
|------------|------------|-------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890  | 1234567890 | 1234567890 |     |
| AATGTTCTC  | CCAAGCCAAA | AATGCCCCCTA | AGATGTATTC | TGGACAATTG | 450 |
| C S S      | Q G K      | N A P K     | M Y S      | G E L      |     |
| N V P P    | K A K      | M P L       | R C I L    | E N W      |     |
| M F L      | P R Q K    | C P .       | D V F      | W R I G    |     |
| GGACCAATCT | GACCCTCAGA | CAGTAAGAAA  | AAAAATGACT | TATATTCTTC | 500 |
| G P I .    | P S D      | S K K       | K N D L    | Y S S      |     |
| D Q S      | D P Q T    | V R K       | K M T      | Y I L L    |     |
| T N L      | T L R      | Q .         | E K K .    | L I F F    |     |
| TGCAGTACCG | CCTGGCCACG | GATATCCTCT  | TCAAGGGGGA | GAAACCTGGC | 550 |
| A V P      | P G H G    | Y P L       | Q G G      | E T W P    |     |
| Q Y R      | L A T      | D I L F     | K G E      | K P G      |     |
| C S T A    | W P R      | I S S       | S R G R    | N L A      |     |
| CTCCTGAGGG | AAGTATAAAT | TATAACACCA  | TCTTACAGCT | AGACCTGTTT | 600 |
| P E G      | S I N      | Y N T I     | L Q L      | D L F      |     |
| L L R E    | V .        | I I T P     | S Y S .    | T C F      |     |
| S .        | G K Y K L  | . H H       | L T A      | R P V L    |     |
| TGTAGAAAAG | GAGGCAAATG | GAGTGAAGTG  | CCATATTTAC | AAACTTTCTT | 650 |
| C R K G    | G K W      | S E V       | P Y L Q    | T F F      |     |
| V E K      | E A N G    | V K C       | H I Y      | K L S F    |     |
| . K R      | R Q M E .  | S A         | I F T      | N F L      |     |
| TTCATTAAAA | GACAACTGGC | AATTATGTAA  | ACAGTGIGAT | TGTGTCTTA  | 700 |
| S L K      | D N S Q    | L C K       | Q C D      | L C P T    |     |
| H .        | K T T R    | N Y V N     | S V I      | C V L      |     |
| F I K R    | Q L A      | I M .       | T V .      | F V S Y    |     |
| CAGGAAGCCC | TCAGATCTAC | CTCCCTACCC  | CGGCATCTCC | CTGACTCCTT | 750 |
| G S P      | Q I Y      | L P T P     | A S P .    | L L        |     |
| Q E A L    | R S T      | S L P       | R H L P    | D S F      |     |
| R K P      | S D L P    | P Y P       | G I S      | L T P S    |     |
| CCCCAACTAA | TAAGGACCCA | CTTCAGCCCA  | AACAGTCCAA | AAGGACATAG | 800 |
| P Q L I    | R T H      | F S P       | N S P K    | G H        |     |
| P N . .    | G P T      | S A Q       | T V Q      | K D I      |     |
| P T N      | K D P      | L Q P K     | Q S K      | R T .      |     |

25 / 32

## FIG 11

| 10         | 20         | 30         | 40          | 50         |     |
|------------|------------|------------|-------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890  | 1234567890 |     |
| GGCATTGATA | GCACCCATCA | GATGGCCAAA | TCATTATTTA  | CTGGACCAGG | 50  |
| G I D S    | T H Q      | M A K      | S L F T     | G P G      |     |
| A L I      | A P I R    | W P N      | H Y L       | L D Q A    |     |
| H . .      | H P S      | D G Q I    | I I Y       | W T R      |     |
|            |            |            |             |            |     |
| CCTTTTCAAA | ACTATCAAGC | AGATAGGGCC | CGTGAAGCAT  | GCCAAAGAAA | 100 |
| L F K      | T I K Q    | I G P      | V K H       | A K E I    |     |
| F S K      | L S S R    | . G P      | . S M       | P K K      |     |
| P F Q N    | Y Q A      | D R A      | R E A C     | Q R N      |     |
|            |            |            |             |            |     |
| TAATCCCTTG | CCTTATCGCC | ATGTTCTTTC | AGGAGAACAA  | AGAACAGGCC | 150 |
| I P C      | L I A      | M F L Q    | E N K       | E Q A      |     |
| . S P A    | L S P      | C S F      | R R T K     | N R P      |     |
| N P L      | P Y R H    | V P S      | G E Q       | R T G H    |     |
|            |            |            |             |            |     |
| ATTACCCAGG | GGAAGACTGG | CAACTAGATT | TTACCCACAT  | GGCCAAATGT | 200 |
| I T Q G    | K T G N    | . I L      | P T W       | P N V      |     |
| L P R      | G R L A    | T R F      | Y P H       | G Q M S    |     |
| Y P G      | E D W      | Q L D F    | T H M       | A K C      |     |
|            |            |            |             |            |     |
| CAGGGATTTC | AGCATCTACT | AGICTGGGCA | GATACITTTCA | CTGGTGGGT  | 250 |
| R D F      | S I Y .    | S G Q      | I L S       | L V G W    |     |
| G I S      | A S T      | S L G R    | Y F H       | W L G      |     |
| Q G F Q    | H L L      | V W A      | D T F T     | G W V      |     |
|            |            |            |             |            |     |
| GGAGICTTCT | CCTTGTAGGA | CAGAAAAGAC | CCAAGAGGTA  | ATAAAGGCAC | 300 |
| S L L      | L V G      | Q K R P    | K R .       | . R H      |     |
| G V F S    | L . D      | R K D      | P R G N     | K G T      |     |
| E S S      | P C R T    | E K T      | Q E V       | I K A L    |     |
|            |            |            |             |            |     |
| TAATGAAATA | ATTCCACAT  | TTGGACTTCC | CCAGGATTA   | CAGGTGACA  | 350 |
| . . N N    | S Q I      | W T S      | P R I T     | G . Q      |     |
| N E I      | I P R F    | G L P      | P G L       | Q G D N    |     |
| M K .      | F P D      | L D F P    | Q D Y       | R V T      |     |

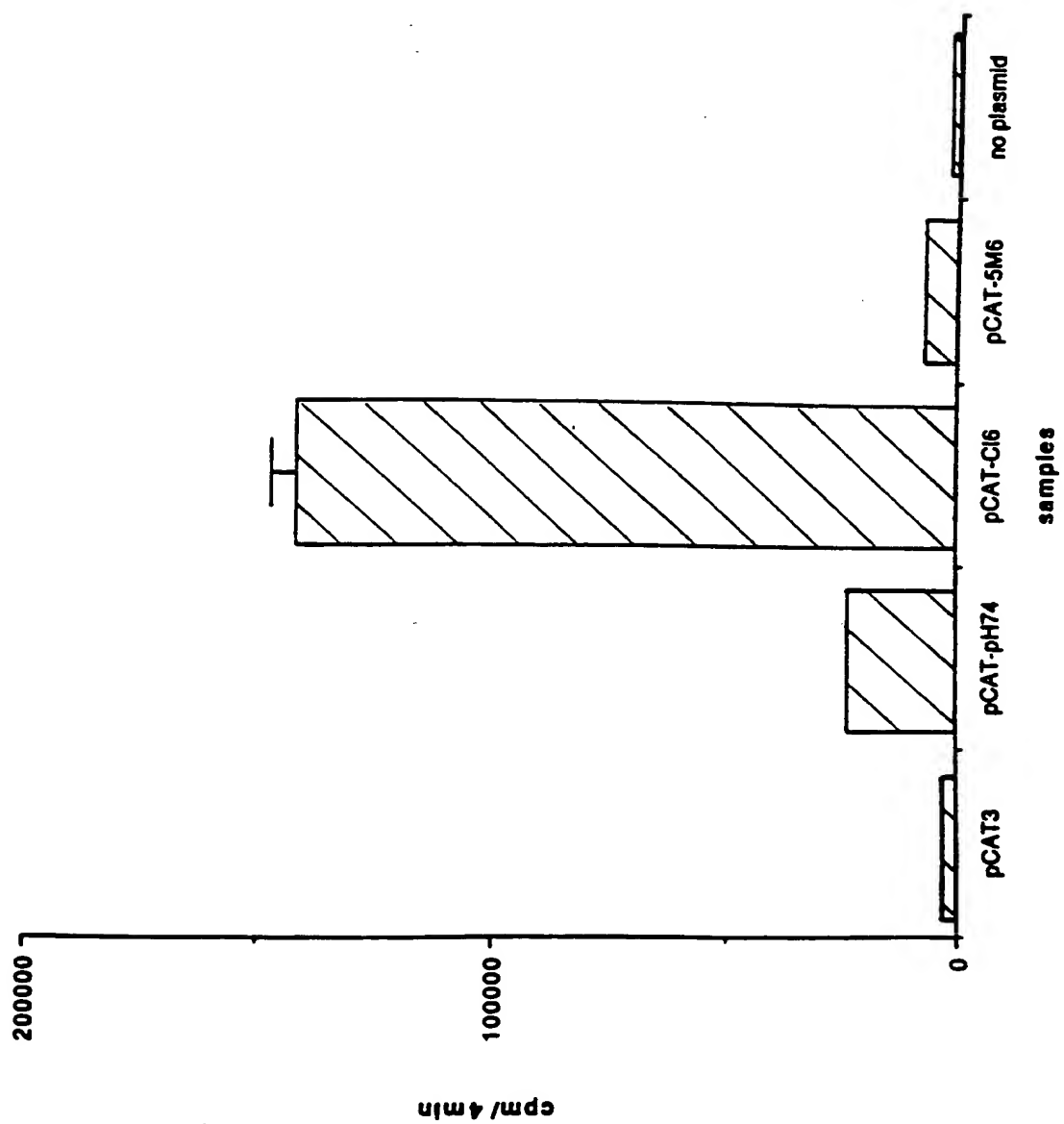


26 / 32  
FIG 11 (continued)

| 10         | 20         | 30         | 40         | 50         |     |
|------------|------------|------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |     |
| ATGGCCCCGC | TTTCAAGGCT | GCAGTAACCC | AGGGAGTATC | CCAGGIGTTA | 400 |
| W P R      | F Q G C    | S N P      | G S I      | P G V R    |     |
| G P A      | F K A      | A V T Q    | G V S      | Q V L      |     |
| M A P L    | S R L      | Q .        | P R E Y P  | R C .      |     |
| GGCATACAAT | ATCACTTACA | CTGTGCCTGG | AGGCCACAAT | CCTCCAGAAA | 450 |
| H T I      | S L T      | L C L E    | A T I      | L Q K      |     |
| G I Q Y    | H L H      | C A W      | R P Q S    | S R K      |     |
| A Y N      | I T Y T    | V P G      | G H N      | P P E K    |     |
| AGTCAAGAAA | ATGAATGAAA | CACTCAAAGA | TCTAAAAAAG | CTAACCCAAG | 500 |
| S Q E N    | E .        | N T Q R    | S K K A    | N P R      |     |
| V K K      | M N E T    | L K D      | L K K      | L T Q E    |     |
| S R K      | . M K      | H S K I    | . K S      | . P K      |     |
| AAACCCACAT | TGCATGACCT | GTTCTGTTGC | CTATAACCTT | ACTAAGAATC | 550 |
| N P H      | C M T C    | S V A      | Y N L      | T K N P    |     |
| T H I      | A .        | P V L L P  | I T L      | L R I      |     |
| K P T L    | H D L      | F C C      | L .        | P Y . E S  |     |
| CATAACTATC | CCCCAAAAG  | CAGGACTTAG | CCCATACGAG | ATGCTATATG | 600 |
| . L S      | P K K      | Q D L A    | H T R      | C Y M      |     |
| H N Y P    | P K S      | R T .      | P I R D    | A I W      |     |
| I T I      | P Q K A    | G L S      | P Y E      | M L Y G    |     |
| GATGGCCTTT | OCTAACCAAT | GACCTTGTGC | TTGACTGAGA | AATGGCCAAC | 650 |
| D G L S    | . P M      | T L C      | L T E K    | W P T      |     |
| M A F      | P N Q .    | P C A      | . L R      | N G Q L    |     |
| W P F      | L T N      | D L V L    | D .        | E M A N    |     |
| TTAGTTGCAG | ACATCACTTC | CTTAGCCAAA | TATCAACAAG | TTCTTAAAAC | 700 |
| . L Q      | T S P P    | . P N      | I N K      | F L K H    |     |
| S C R      | H H L      | L S Q I    | S T S      | S . N      |     |
| L V A D    | I T S      | L A K      | Y Q Q V    | L K T      |     |

27 / 32  
FIG 11 (continued)

| 10                                                      | 20         | 30         | 40         | 50         |     |
|---------------------------------------------------------|------------|------------|------------|------------|-----|
| 1234567890                                              | 1234567890 | 1234567890 | 1234567890 | 1234567890 |     |
| ATCACAGGGA ACCTGTCCCC GAGAGGAGGG AAAGGAACTA TTCCACCCCTG |            |            |            |            | 750 |
| H R E P V P E R R E R N Y S T L                         |            |            |            |            |     |
| I T G N L S P R G G K G T I P P W                       |            |            |            |            |     |
| S Q G T C P R E E G K E L F H P G                       |            |            |            |            |     |
| GIGACATG                                                |            |            |            |            | 758 |
| V T                                                     |            |            |            |            |     |
| . H                                                     |            |            |            |            |     |
| D M                                                     |            |            |            |            |     |



|      |                                                                                                              |
|------|--------------------------------------------------------------------------------------------------------------|
| 100  | ATGGCCCTCC CTATATGAC TTCTCTCTT ACTCTCTCT TACCCCTTT GCTCTCTCT GACCCCTC CATCTCTCT TACACGAGT AGCTCCCTT          |
| 34   | MA L P Y M T P L P I V L L P P P A L T A P P P C C C T T T B S S P Y                                         |
|      | signal peptide                                                                                               |
| 200  | ACCAAGATT TCTATGAGA AGCCGCTTC CTGGAAATAT TATGCGCCA TCATATGGA GTTATCTTA GCGAACTCT ACCTCTACTO CCAACACCA        |
| 67   | Q E P L . R T R L P G N I D A P S Y R S L S K G N S I P T A H T H                                            |
| 300  | TATGCCGCC AACTGCTATA ACTCTGCCAC TCTTTCATG CATGCAATA CTCTATCT GACAGGAAA ATGATTATCT CTAGTCTCTC TGGAGGACTT      |
| 100  | H P R M C Y M S A T L C M H A N T H Y W T O K M I N P S C P O O L                                            |
| 400  | GGAGCCACTG TCTCTTGAC CTATCCAGTA TGTCTATG GGTGTGATT GAGGTCTAG CAGAGAAA ACAAGTAAAG GAGGAAATCT                  |
| 134  | Q A T V C M T Y P T H T S M S D G G I Q G O A R E K Q V K E A I S                                            |
| 500  | CCCACTGAC CGGGAGCAT AGCAGCCCTA GCGCTACAA AGCACTAGTT CTCTCAAAC TACATGAACT CTCTCTGAC CATCTCTGCC TGGTGAAGCT     |
| 167  | Q L T R G H S T P E P Y K G L V L S K L H E T L R T H T R L V S L                                            |
| 600  | ATTATATCC AGCTCTACT GAGTCCAGTA GAGTCTAGCC CAATACTTA CTAACTCTT GATCTCTCT CCGCTGACT TCAGCCGATA CATTCATCT       |
| 200  | P N T T L T R L M E V S A Q N P T N C M H C L P L H P R P Y I S I                                            |
| 700  | CTCTTCTG ACAGATGGA CAAGTCTGAC ACAGATTA ACAGCTCTT GCTTTGATA GAGCTCTG TTTCGAACT GTTCTAACT GGAATGAACT CATACCTGA |
| 234  | P V P E Q M M H P S T E I N T S V L V Q P L V S N L E I T H T S N                                            |
| 800  | ACCTGCTG TGTAAATTT AGCAATCTA TAGACGAC CAGCTCCAA TGCATGAGT GGTATGAC TCCTCAAGCA ATAGCTGCC TACCTCTAGG           |
| 267  | L T C V K P S M T I D T T S S Q C I R M V T P P T R I V C L P S G                                            |
| 900  | ATAATTTT GCTCTGATA CTCTAGCTA TCAATCTT AGGCTCTT CAGATCTAT GCTCTCTCT TCACTCTTGG TCCCTCTAT GACCATCTAC           |
| 300  | I F P V C G T S A Y H C L N G S S E S M C F L S F L V P P M T I Y                                            |
| 1000 | ACTGACAG ATTATCTA TCATCTGTA CTTAGCCCC ACACAAAG AGTACCAAT CTCTCTTGG TTATCAGAG AGGATCTTA GCGAGCTAG             |
| 334  | T E Q D L Y M H V V P K P H N K R V P I L P P V I R A O V L G R L G                                          |
| 1100 | GTACTGCAAT TGGAGTATC ACAGCTCTA CTGAGTCTA CTCAATCTA TCTGAGAAA TAATGTGA GTGAGAGG GTACTGACT CCGTGTGAC           |
| 367  | T G I G S I T T S T Q F Y Y K L S Q E I N G D H E Q V T D S L V T                                            |
| 1200 | CTTGCAAGT CAAGTAACT GCGTAGGAG AGTAGCTTT CAATCTGA GAGCTTGA CTCTCTACC GCAAAAGG GCGGAACTO TTATTTTA              |
| 400  | L Q D Q L N S L A A V V L Q N R R A L D L L T A K R G G T C L P L                                            |
| 1300 | GGAGAGAC GCTGTATTA TGTAACTGA TCACTGAGAA AGTTAAGAA ATTGAGATC GATATCAATG TAGAGAGAG GAGCTTCAA                   |
| 434  | G E R C Y Y V N U S R I V T E K V K E I R D R I Q C R A E E L Q N                                            |
| 1400 | ACAGCAAG CTGGGCCCTC CTGAGCAAT GATGCCCTG GGTCTCCCT TTCTAGGAC CTCTAGGAC TCTAATATG TTACTCTCT TGGAGCCCTG         |
| 467  | T E R M G L L S Q M H P M V L P F L G P L A A L I L L L L L P Q P C                                          |
| 1500 | TATCTTAACT CTCTCTGTA AGTTCTCTC TTCCAGATTT GAGCTGTA AGCTAGAT GTCTTACAA ATGCACTCC AGATGAGTC CATGCTAGS          |
| 500  | I F N L L V K P V S S R I E A V K L Q M V L Q H E P Q H E S H T K                                            |
| 1600 | ATCCAGCTG GAGCCCTGGA CGGCCCTGCT AGCCATGCT CCGATGTA TGAATTTAA GGCACCCCTC CCGAGAAAT CTCACTGCA CACCCCTAC        |
| 534  | I H R G P L D R P A S P C S D V N D I E G T P P E E I S T A Q P L L                                          |
| 1700 | TATGCCCTAA TTGAGCGGAA AGCATTTAG CCGCTCATCA GCGACCTCC CCAACAGAC TTGGTTTTT CTGTGTAGAG GCGGAGCTCA GAGACAGAC     |
| 542  | C P N S A O S S                                                                                              |
| 1800 | TAGCTGGATT TGTAGGCA ACCAGATCT CTTAGCTTA GCTGGAGG TGACTGATC CAGCTTAA CATGGGCTT GCAGTATG TCACACCGCA            |
| 1900 | CCATCAGAG AGCTGCTAA ATGCTGATTT AGGCAATAT AGAGGTAA GAATAGCA ATCATCTATT GCTTGAGAG ACAGCGGAG GACAGAGAT          |
| 2000 | CGGATATAA ACCAGGCTT TCGAGCGGC ACCGCACT CCGTTGGCT CCGCTGCTT TGTATGGCG CTCTGTTTTT ACTCTATTTT ACTCTATTA         |
| 2010 | ATCTTGAC TGAAGAAA AAAAAAAAAA                                                                                 |
|      | Cap site                                                                                                     |
|      | Poly A signal                                                                                                |

FIG13

FIG 14

CAGCAACCCC CTTTGGGTCC CTTCCCATTTG TATGGGAGCT CTGTTTTCAC TCTATTATTA CATGCAACTG CACTCTCTG GTCCGCTGTTT  
 TTTATGGCTC AAGCTGAGCT TTTGTTCCGC ATCCACCACT GCTGTTTCCG ACCGCTGCT GACTTCCATC CCTTGGATC CAGCAGATG  
 TCCGCTGTGC TCTGTATCCA GCACAGGCGC CCATGGCTC TCCCAATGG GCTAAAGGCT TGCCATTTGT CCTGCACAGC TAAGTGCTG GGTTCATCCT  
 AATCAGCTG AACACTAGTC ACTGGGTCC ACGGTCTCT TCCATGACCC ATGGCTTCTA ATAGAGCTAT AACACTCACT GCATGGTCCA AGATTCATTT  
 CCTTGGAAATC CGTGAGACCA AGAACCCCG GTGAGAGAAC ACNAGGCTG CCACCATGTT GGAAGCAGCC CACCACCATTT TTGGAAGCAG CCGGCCACTA  
 TCTTGGGAGC TCTGGGAGCA AGGACCCCG GTAAACAATTT GGTGACACG AAGGACCTG AATCCGCCAC CATGAAGGGA TCTCCAAAGC ATGGGAAAC  
 GTTCCCCCG AGGCAAAAT GGCCTTAGAA CGTATTCTGG AGAATGGGA CCAATGTGAC ACTCAGACGC TARGAAGAA ACGATTATTA TTCTCTGCA  
 V P P E A K M P L E R I L E N W D Q C D T Q T L R K K R F I F F C S  
 GTACGGCTG GCCACATAT CTTCTTCAG GGAGAGAAC CTGGCTTCTT GAGGNAAGTA TAAATTTATA CATCATCTTA CAGCTAGACC TCTTCTGTAG  
 T A W P Q Y P L Q Q R E T W L P E G S I N Y N I I L Q L D L F C R  
 AAGGAGGC AATGGAGTG AAGTGCCATA TGTGCAACT TCTTTTCAT TAAGAGACAA CTCACAATTA TGTAAAAAGT GTGGTTTATG CCTACAGGA  
 K E G K W S E V P Y V Q T F F S L R D N S Q L C K K C G L C P T G  
 AGCCTCAGA GTCCACCTCC CTACCCCGC GTCCCTCTTC CTCACTAAT AAGGACCCCTT CTTTAACCCA AAGGTCCTCA AAGGAGATAG  
 S P Q S P P P Y P S V P S P T P S S T N K D P P L T Q T V Q K E I D  
 ACAAGGGGT AAACAATGAA CCAAGAGTG CCAATATTC CCGATTATGC CCCCTCCAG CAGTGAGAGG AGGAGAATTC GGCACAGCA GAGTGCTGT  
 K G V N N E P K S A N I P R L C P L Q A V R G G E F G P A R V P V  
 ACCTTTTCT CTCTCAGACT TAAAGCAAT TAAATAGAC CTAGGTAAT TCTCAGATAA CCTGACGGC TATATTGATG TTTTACAAGG GTTAGGACAA  
 P F S L S D L K Q I K I D L G K F S D N P D G Y I D V L Q G L G Q  
 TCTTTGATC TGACATGGAG AGATATATG TTAATACTAA ATCAGACACT AACCCCAAT GAGAGAAGT CCGCTGTAC TGCAGCCCA GAGTTGCGG  
 S F D L T W R D I M L L L N Q T L T P N E R S A A V T A A R E F G D  
 ATCTTTGTA TCTCAGTCAG GCCACAATA GGATGACAC AGAGGAAGA ACAACTCCA CAGGCCAGCA GGCAGTTCCC AGTGTAGACC CTCAITGGGA  
 L W Y L S Q A N N R M T T E E R T P T G Q Q A V P S V D P H W D  
 CACAGAATCA GAACATGGAG ATTGGTGCCA CAACATTTG CTAACTTGG TGCTAGAAG ACTAGGAAG AGCCTATGAA TTACTCAATG  
 T E S E H G D W C H K H L L T C V L E G L R K T R K K P M N Y S M  
 ATGTCCACTA TAACACAGG AAAGGAAGAA AATCTTACTG CTTTCTGGA CAGACTAAG GAGGATGGA GGAAGCATAC CTCCCTGTCA CCTGACTCTA  
 M S T I T Q G K E E N L T A F L D R L R E A L R K H T  
 TTGAAGGCA ACTAATCTTA AAGGATAGT TTAATCTCA GTACATGCA GACATTAGAA AAAACTTCA AAGTCCGTC TTAGGCTCGG AACAAAACIT  
 E G Q L I L K D K F I T Q S A A D I R K K L Q K S V L G S E Q N L  
 AGAACCTTA TTGAATCTG CAACCTCGT TTTTATANT AGAGATCAGG AGGACAGGC AGATGGAG AATGGGATA AAAAAAAG GGCACCGCT  
 E T L L N L A T S V F Y N R D Q E E Q A E W D K W D K K R A T A  
 TTAGTCATGG CCTCAGGCA AGCGGACTTT GGAGCTCTG GAAAGGAA AGCTGGGA ANTAGAAGC CTAATAGGC TTGCTCCAG TGCCTCTAC  
 L V M A L R Q A D F G G S G K G K S W / A N R K P N R A C F Q C G L Q  
 AAGGACACTT TAAAAAGAT TGTCCAATA GAATAAGCC GCCCTTGT CCGTCCCTT TACGCTCAGG GAATCACTG AAGGCCACT GCCCAGGGG  
 G H F K K D C P N R N K P P P C / R P C P L R Q G N / H W K A H C P R G  
 ATCAAGATAC TCTGAGTCAG AAGCCATTA CCAGATGATC CAGCAGCAGG ACTGA  
 S R Y S E S E A I N Q M I Q Q Q D

## FIG 15

GGACCCGTAG TATGGGGTAA TCCCTCTCCG GAAACCAAGC CCCAGTACTC AGAAGAAGAA ATAGAATGGG GAACCTCAGC AGGACATGCT TTCTCTCCCT  
 100 G | P V V W G N P L R E T K P Q Y S E E E I E W G T S R G H G F L P S 34  
 CAGGATGGCT AGCCACTGAA GAAGGAAGAA TACTTTTCT GGCAGCTAAC CAATGGAAAT TACTTAAAC CCTTCAGCA ACCTTCCACT TAGGCATTGA  
 200 G W L A T E E G K I L L L A A N Q W K L L K T L Q Q T F L L G I D 67  
 TAGCACCCAT CAGATAGCCA ATCATTTATT TACTGGACCA GGCCTTTTCA AACTATCA GCAGATAGTC AGGCTCTGTG AAGTGTGCCA AAGAATATAT  
 300 S T E Q I A K S L F T G P G L P K T I K Q I V R A E E V E Q R N N 100  
 CCCCTGCCCT ATGCCAAGC TCCTTCAGGA GAACAAAGAA CAGGCATTA CCCAAGAGAA GACTGGCAAC TAGATTTTAT CCACATGCCA AAATCACAAG  
 400 P L P Y R Q A P S G E Q R T G N Y P R E D W Q L D F I H M P K S Q G 134  
 GATTTCAGTG TCTACTAGTC TGGTAGATA CTTTCACTGG TTGGGACAG GCTTCTCCCT GTAGGACAGA AAGTTCGA GAGGTATATA AGGCACTAGT  
 500 F Q C L L Y W V D T F T G W A E A P P C R T E K P Q E V I K A L V 167  
 TCATGAAGTA ATTCCAGAT TGGACTTCC CTGAGGCTTA CAGAGTGACA ATGCTCTGC TTTCAGGCC ACAGTACCC AGGGAGTATC CCAGGCGTTA  
 600 H E V I P R P G L P . G L Q S D N G P A F K A T V T Q G V S Q A L 200  
 GGTATAGAAT ATCACTTACA CTGCACCTAG AGCCACAAAT CCTCAGGGA GGTTCAGAA ATCA~~AA~~ACAC TCAAGGACA TCTAACAAG CTAAACCCAGG  
 700 G I E Y H L H C T . R P Q S S G K V E K M K T L K R H L N K L T Q E 234  
 AAACCCACT GGCATGGTCT GCTCTGTGT CTATAGCCTT ACTAGAAATC CAAACTCTC CCCAAGGCC AGGACTTAGC CCATACAGAA TCGTGTATGG  
 800 T H L A W S A L L S I A L L R I Q N S P Q K A G L S P Y R M L Y Q 267  
 ACGTCTCTTC CTAAACCAATG ACCTTCTGCT TCACCAAGAG ATGGCCAACT TAGTTCAGCA CATCACTCC TTAGCCAAAT ATCAACAAGT TCTTAAACA  
 900 R S F L T W D L L L D Q E M A N L V A D I T S L A K Y Q Q V L K T 300  
 TTACAAAGAG CCTGTCCCG AGAGGAGGA AAGAATAT TCCACCTGG TGTATGTA TTACTCAAGT CCTTCCCTC TAATCCCA TCCCTAGACA  
 1000 L Q G A C P R E E G K E I F H P G V M V L V K S L P S M S P S L D T 334  
 CATCTGGGG AGGACCTTAC CCAGTCATTT TATCTATCC AACTGCGGT AAGTGGCTG GAGTGGAGTC TTGATACAT CAACTCGAA TCAAACTCG  
 1100 S W G G P Y P V I L S I P T A V K V A G V E S W I H H T R I K P W 367  
 GATACTCCG AAGGAACCCG AAATCCAGG GGACACGCT AGCTATTCT TTGAACCTT AGAGATCTG TCGTCTCT TCAAGCAACA ACGTGA  
 1197 I L P K E P E N P G D N A S Y F F E P L E D L C L L P K Q Q P 398

100 GAGAGGCA GATATATG GCTGGGCA GAGGAGAG ACAGAGCA GTATAGAA AAGAGGAA GTAGAGAA GAAAGAGA GAGAGGAA  
 E N S S I S W L A E V G K D S K K . R K K G E S Q R K K K R E E E T  
 200 QAGAGGAA GTAGAGCA GAGAGGAA GTATAGAA AAGAGGAA GCTATGCTT TAAAGGAG GTAAATTC TGCTACTA GCAAGGCAAT  
 K K N L K R E S S K E K T V Y P I P L K A R V N F C L P S Q G I  
 300 ATCTCTTA TGTGAAAT GACCTGAT CTTCTGAC ACTATGCA CAGGAGAG AAGCTGTC TTCTATATC GCAATATTA CATTGAGCA  
 F P L C G T S T Y I C L P T N W T G T R T L V F L S P N I N I A P  
 400 GAAATGCA GCTATGCTT ACTATGCA GCTATGTC GTATGAG ACCAGAGCA CTAAATGCTT TATTGAGG GTATGAGAG GCTACTCTA  
 G N Q T L L V P V K A K V R Q C R A I Q L I S L F I G L G M A T A T  
 500 GAGAGGAG AATAGGCTT TGTCTGCT GATATGCTA CTAAATGCA ACTATGCA CATTGAGCA GAAATATTA AATCTATCT  
 G T G I A G L S T S L S Y Y H T L S K N F S D S L Q E I H K S I L  
 600 TACTATCA TGTGAGAG ACTATGCTT ACTATGCTT GCTATGAG GCTATGCTT ACTATGCTT ACTATGCTT ACTATGCTT  
 T L Q S Q L D S L A A H T L Q N R R G P H L L T A E K G G L C T P  
 700 TAAAGGAG AATAGGCTT TGTCTGCT GATATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA  
 L G E C C F Y T N Q S G I V R D A T W H L Q E R A S D I R Q C L S  
 800 GAAATGCA TGTATGCTT GAAATGCTT TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA  
 N S Y T N L M S W A T W L L P F L G P H A A I L L L L T F G P C I  
 900 TGTATGCTT GATATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA  
 F K L L V K F V S S R I E A I K L Q M V L Q H E P Q M S S T N N F  
 1000 TAAAGGAG AATAGGCTT TGTCTGCT GATATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA  
 Y Q G P L E R S T G T S T S L E I P L W K T L Q L Q G P F F A P I Q  
 1100 AAGAGGAG AATAGGCTT TGTCTGCT GATATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA  
 Q E V A R A V I G Q I P N S S W G V L F R G G I E E . A C W Q P  
 1200 TGTATGCTT GATATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA  
 H S P R W I S V P P Q P W C P L W P C L R S P S A C H C T V G A S  
 1300 TGTATGCTT GATATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA  
 F W A G Q G R S Q L P Q L A G R Y G G R D A G G N Q G C A W R L R A  
 1400 GAGAGGAG AATAGGCTT TGTCTGCT GATATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA  
 S H S S R W A W A R A P H S G S E G L S T W A R Q M L C S T S S  
 1500 GATATGCTT GATATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA  
 L G L S C L P R G A G L R E H A A C P C L S P P P R R G F L H S P  
 1600 AATATGCTT GATATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA  
 S F P D K H H P L S T V P S P I N H P R V E E C G H T A R D W Q A V  
 1700 TGTATGCTT GATATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA  
 P L A A L V R D P L R E A S W A P E S G G D L E N L Y V L R D C  
 TAAAGGAG AATAGGCTT TGTCTGCT GATATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA TGTATGCTA  
 K Y T N Q H

1719

FIG 16



## DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIÈRE DE BREVETS (PCT)

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                      |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------|
| (51) Classification internationale des brevets <sup>6</sup> :<br><b>C12N 15/00</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | <b>A2</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | (11) Numéro de publication internationale: <b>WO 99/02666</b><br>(43) Date de publication internationale: 21 janvier 1999 (21.01.99) |
| <p>(21) Numéro de la demande internationale: <b>PCT/FR98/01460</b></p> <p>(22) Date de dépôt international: 7 juillet 1998 (07.07.98)</p> <p>(30) Données relatives à la priorité:<br/>97/08816 7 juillet 1997 (07.07.97) <b>FR</b></p> <p>(71) Déposant (pour tous les Etats désignés sauf US): <b>BIO MERIEUX [FR/FR];</b> Chemin de l'Orme, F-69280 Marcy l'Etoile (FR).</p> <p>(72) Inventeurs; et</p> <p>(75) Inventeurs/Déposants (US seulement): <b>PARAH-NOS-BACCALA, Glauca [FR/FR];</b> 75, cours Duguesclin, F-69003 Lyon (FR). <b>KOMURIAN-PRADEL, Florence [FR/FR];</b> 114, chemin du Pavillon, F-69250 Poleymieux au Mont d'Or (FR). <b>BEDIN, Frédéric [FR/FR];</b> 6, rue Gaspard André, F-69002 Lyon (FR). <b>SODOYER, Mireille [FR/FR];</b> 5, rue du Brûlet, F-69110 Sainte Foy les Lyon (FR). <b>OTT, Catherine [FR/FR];</b> 103, avenue Berthelot, F-69007 Lyon (FR). <b>MALLET, François [FR/FR];</b> 84, rue Anatole France, F-69100 Villeurbanne (FR). <b>PERRON, Hervé [FR/FR];</b> 15, rue de Boyer, F-69005 Lyon (FR). <b>MANDRAND, Bernard [FR/FR];</b> 21, rue de la Doua, F-69100 Villeurbanne (FR).</p> | <p>(74) Mandataire: <b>CABINET GERMAIN &amp; MAUREAU;</b> Boîte postale 6153, F-69466 Lyon Cedex 06 (FR).</p> <p>(81) Etats désignés: <b>AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, brevet ARIPO (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), brevet eurasién (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</b></p> <p><b>Publiée</b><br/><i>Sans rapport de recherche internationale, sera republiée dès réception de ce rapport.</i></p> |                                                                                                                                      |
| <p>(54) Title: <b>RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES</b></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                      |
| <p>(54) Titre: <b>MATERIEL NUCLEIQUE RETROVIRAL ET FRAGMENTS NUCLEOTIDIQUES NOTAMMENT ASSOCIES A LA SCLEROSE EN PLAQUES-ET/OU LA POLYARTHRITE RHUMATOIDE, A DES FINS DE DIAGNOSTIC, PROPHYLACTIQUES ET THERAPEUTIQUES</b></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                      |
| <p>(57) Abstract</p> <p>The invention concerns a nucleic material, in isolated or purified state, and a nucleotide fragment comprising a nucleotide sequence selected from the group consisting in (i) the sequences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 and SEQ ID NO: 142, (ii) the complementary sequences of sequences (i); and (iii) the sequences equivalent to sequences (ii) and (iii), in particular the sequence having for every series of 100 contiguous monomers, at least 50 %, preferably 70 % homology with sequences (i) and (ii) respectively. The invention also concerns their uses for detecting a retrovirus associated with multiple sclerosis and/or rheumatoid arthritis.</p>                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                      |
| <p>(57) Abrégé</p> <p>Matériel nucléique, à l'état isolé ou purifié, et fragment nucléotidique, comprenant une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 et SEQ ID NO: 142, (ii) les séquences complémentaires des séquences (i); et (iii) les séquences équivalentes aux séquences (i) ou (ii), en particulier les séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins 70 % d'homologie avec respectivement les séquences (i) ou (ii), et utilisations pour détecter un rétrovirus associé à la sclérose en plaques et/ou la polyarthrite rhumatoïde.</p>                                                                                                                                                                                                                                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |                                                                                                                                      |



# **UNIQUEMENT A TITRE D'INFORMATION**

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

|    |                           |    |                                               |    |                                          |    |                       |
|----|---------------------------|----|-----------------------------------------------|----|------------------------------------------|----|-----------------------|
| AL | Albanie                   | ES | Espagne                                       | LS | Lesotho                                  | SI | Slovénie              |
| AM | Arménie                   | FI | Finlande                                      | LT | Lituanie                                 | SK | Slovaquie             |
| AT | Autriche                  | FR | France                                        | LU | Luxembourg                               | SN | Sénégal               |
| AU | Australie                 | GA | Gabon                                         | LV | Lettonie                                 | SZ | Swaziland             |
| AZ | Azerbaïdjan               | GB | Royaume-Uni                                   | MC | Monaco                                   | TD | Tchad                 |
| BA | Bosnie-Herzégovine        | GE | Géorgie                                       | MD | République de Moldova                    | TG | Togo                  |
| BB | Barbade                   | GH | Ghana                                         | MG | Madagascar                               | TJ | Tadjikistan           |
| BE | Belgique                  | GN | Guinée                                        | MK | Ex-République yougoslave<br>de Macédoine | TM | Turkménistan          |
| BF | Burkina Faso              | GR | Grèce                                         | ML | Mali                                     | TR | Turquie               |
| BG | Bulgarie                  | HU | Hongrie                                       | MN | Mongolie                                 | TT | Trinité-et-Tobago     |
| BJ | Bénin                     | IE | Irlande                                       | MR | Mauritanie                               | UA | Ukraine               |
| BR | Bésil                     | IL | Israël                                        | MW | Malawi                                   | UG | Ouganda               |
| BY | Bélarus                   | IS | Islande                                       | MX | Mexique                                  | US | Etats-Unis d'Amérique |
| CA | Canada                    | IT | Italie                                        | NE | Niger                                    | UZ | Ouzbékistan           |
| CF | République centrafricaine | JP | Japon                                         | NL | Pays-Bas                                 | VN | Viet Nam              |
| CG | Congo                     | KE | Kenya                                         | NO | Norvège                                  | YU | Yougoslavie           |
| CH | Suisse                    | KG | Kirghizistan                                  | NZ | Nouvelle-Zélande                         | ZW | Zimbabwe              |
| CI | Côte d'Ivoire             | KP | République populaire<br>démocratique de Corée | PL | Pologne                                  |    |                       |
| CM | Cameroun                  | KR | République de Corée                           | PT | Portugal                                 |    |                       |
| CN | Chine                     | KZ | Kazakstan                                     | RO | Roumanie                                 |    |                       |
| CU | Cuba                      | LC | Sainte-Lucie                                  | RU | Fédération de Russie                     |    |                       |
| CZ | République tchèque        | LI | Liechtenstein                                 | SD | Soudan                                   |    |                       |
| DE | Allemagne                 | LK | Sri Lanka                                     | SE | Suède                                    |    |                       |
| DK | Danemark                  | LR | Libéria                                       | SG | Singapour                                |    |                       |
| EE | Estonie                   |    |                                               |    |                                          |    |                       |

**MATERIEL NUCLEIQUE RETROVIRAL ET FRAGMENTS NUCLEOTIDIQUES  
NOTAMMENT ASSOCIES A LA SCLEROSE EN PLAQUES ET/OU LA  
POLYARTHRITE RHUMATOIDE, A DES FINS DE DIAGNOSTIC,  
PROPHYLACTIQUES ET THERAPEUTIQUES**

5 La Sclérose en Plaques (SEP) est une maladie démyélinisante du système nerveux central (SNC) dont la cause complète reste encore inconnue.

De nombreux travaux ont étayé l'hypothèse d'une étiologie virale de la maladie, mais aucun des virus  
10 connus testés ne s'est avéré être l'agent causal recherché: une revue des virus recherchés depuis des années dans la SEP a été faite par E. Norrby et R.T. Johnson.

Récemment, un rétrovirus, différent des rétrovirus  
15 humains connus a été isolé chez des patients atteints de SEP. Les auteurs ont aussi pu montrer que ce rétrovirus pouvait être transmis in vitro, que des patients atteints de SEP produisaient des anticorps susceptibles de reconnaître des protéines associées à l'infection des  
20 cellules leptoméningées par ce rétrovirus, et que l'expression de ce dernier pouvait être fortement stimulée par les gènes immédiats-précoces de certains herpesvirus.

Tous ces résultats plaident en faveur du rôle dans  
la SEP d'au moins un rétrovirus inconnu ou d'un virus  
25 ayant une activité transcriptase inverse (RT) détectable selon la méthode publiée par H. Perron et qualifiée d'activité "RT de type LM7".

Les travaux de la Demanderesse ont permis  
d'obtenir deux lignées continues de cellules infectées par  
30 des isolats naturels provenant de deux patients différents atteints de SEP, par un procédé de culture tel que décrit dans le document WO-A-93 20188, dont le contenu est incorporé par référence à la présente description. Ces deux lignées dérivées de cellules de plexus-choroïdes  
35 humains, dénommées LM7PC et PLI-2 ont été déposées à l'E.C.A.C.C. respectivement le 22 juillet 1992 et le 8

janvier 1993, sous les numéros 92 072201 et 93 010817, conformément aux dispositions du Traité de Budapest. Par ailleurs, les isolats viraux possédant une activité RT de type LM7 ont également été déposés à l'E.C.A.C.C. sous la  
5 dénomination globale de "souches". La "souche" ou isolat hébergé par la lignée PLI-2, dénommée POL-2, a été déposée auprès de l'E.C.A.C.C. le 22 juillet 1992 sous le n° V92072202. La "souche" ou isolat hébergé par la lignée LM7PC, dénommée MS7PG, a été déposée auprès de  
10 l'E.C.A.C.C. le 8 janvier 1993 sous le n° V93010816.

A partir des cultures et des isolats précités, caractérisés par des critères biologiques et morphologiques, on s'est ensuite attaché à caractériser le matériel nucléique associé aux particules virales  
15 produites dans ces cultures.

Les portions de génome déjà caractérisées ont été utilisées pour mettre au point des tests de détection moléculaire du génome viral et des tests immunosérologiques, utilisant les séquences d'acides  
20 codées par les séquences nucléotidiques du génome viral, pour détecter la réponse immunitaire dirigée contre des épitopes associés à l'infection et/ou l'expression virale.

Ces outils ont déjà permis de confirmer une association entre la SEP et l'expression des séquences  
25 identifiées dans les brevets cités plus loin. Cependant, le système viral découvert par la Demanderesse, s'apparente à un système rétroviral complexe. En effet, les séquences retrouvées encapsidées dans les particules virales extracellulaires produites par les différentes  
30 cultures de cellules de patients atteints de SEP, montrent clairement qu'il y a co-encapsidation de génomes rétroviraux apparentés, mais différents du génome rétroviral "sauvage" qui produit les particules virales infectantes. Ce phénomène a été observé entre des  
35 rétrovirus réplicatifs et des rétrovirus endogènes appartenant à la même famille, voire même hétérologues. La

notion de rétrovirus endogène est très importante dans le contexte de notre découverte car, dans le cas de MSRV-1, on a observé que des séquences rétrovirales endogènes comprenant des séquences homologues au génome MSRV-1, existent dans l'ADN humain normal. L'existence d'éléments rétroviraux endogènes (ERV) apparentés à MSRV-1 par tout ou partie de leur génome, explique le fait que l'expression du rétrovirus MSRV-1 dans les cellules humaines puisse interagir avec des séquences endogènes proches. Ces interactions sont retrouvées dans le cas de rétrovirus endogènes pathogènes et/ou infectieux (par exemple certaines souches écotropes du Murine Leukaemia virus), dans le cas de rétrovirus exogènes dont la séquence nucléotidique peut être retrouvée partiellement ou en totalité, sous forme d'ERVs, dans le génome de l'animal hôte (ex. virus exogène de la tumeur mammaire de la souris transmis par le lait). Ces interactions consistent principalement en (i) une transactivation ou co-activation d'ERVs par le rétrovirus répliquatif, (ii) une encapsidation "illégitime" d'ARN apparentés d'ERVs, ou d'ERVs -voire d'ARN cellulaires- possédant simplement des séquences d'encapsidation compatibles, dans les particules rétrovirales produites par l'expression de la souche répliquative, parfois transmissibles et parfois avec une pathogénicité propre, et (iii) des recombinaisons plus ou moins importantes entre les génomes co-encapsidés, notamment dans les phases de transcription inverse, qui conduisent à la formation de génomes hybrides, parfois transmissibles et parfois avec une pathogénicité propre.

Ainsi, (i) différentes séquences apparentées à MSRV-1 ont été retrouvées dans les particules virales purifiées; (ii) l'analyse moléculaire des différentes régions du génome rétroviral MSRV-1 doit être faite en analysant systématiquement les séquences co-encapsidées, interférantes et/ou recombinées qui sont générées par l'infection et/ou l'expression de MSRV-1, de plus,

certaines clones peuvent avoir des parties de séquences défectives produites par la réplication rétrovirale et les erreurs de matrice et/ou de transcription de la transcriptase inverse; (iii) les familles de séquences apparentées à une même région génomique rétrovirale sont les supports d'une détection diagnostique globale qui peut être optimisée par l'identification de régions invariables parmi les clones exprimés et par l'identification de trames de lectures responsables de la production de polypeptides antigéniques et/ou pathogènes qui peuvent n'être produits que par une partie, voire un seul, des clones exprimés et dans ces conditions, l'analyse systématique des clones exprimés dans une région d'un gène donné permet d'évaluer la fréquence de variation et/ou de recombinaison du génome MSRV-1 dans cette région et de définir les séquences optimales pour les applications, notamment diagnostiques; (iv) la pathologie provoquée par un rétrovirus tel que MSRV-1 peut être un effet direct de son expression et des protéines ou peptides produits de ce fait, mais aussi un effet de l'activation, de l'encapsidation, de la recombinaison de génomes apparentés ou hétérologues et des protéines ou peptides produits de ces faits ; ainsi ces génomes associés à l'expression de et/ou l'infection par MSRV-1 sont-ils une partie intégrante de la pathogénicité potentielle de ce virus et donc, constituent des supports de détection diagnostique et des cibles thérapeutiques particulières. De même, tout agent associé à, ou, co-facteur de ces interactions responsables de la pathogénie en cause, tel que MSRV-2 ou le facteur gliotoxique décrit dans la demande de brevet publiée sous le N° FR-2 716 198, peut participer à l'élaboration d'une stratégie globale et très efficace de diagnostic, de pronostic, de suivi thérapeutique et/ou de thérapeutique intégrée de la SEP notamment, mais aussi de toute autre maladie associée aux mêmes agents.

Dans ce contexte, on a fait une découverte parallèle dans une autre maladie autoimmune, la polyarthrite rhumatoïde (PR), qui a été décrite dans la demande de brevet français déposée sous le N°95 02960.

5 Cette découverte montre que, en appliquant des approches méthodologiques similaires à celles qui furent utilisées dans les travaux de la Demanderesse sur la SEP, on a pu identifier un rétrovirus exprimé dans la PR qui partage les séquences décrites pour MSRV-1 dans la SEP et aussi,

10 la co-existence d'une séquence associée MSRV-2 également décrite dans la SEP. En ce qui concerne MSRV-1, les séquences détectées communément dans la SEP et la PR, concernent les gènes *pol* et *gag*. En l'état actuel des connaissances, on peut associer les séquences *gag* et *pol*

15 décrites aux souches MSRV-1 exprimées dans ces deux maladies.

La présente demande de brevet a pour objet différents résultats, supplémentaires par rapport à ceux déjà protégés par les demandes de brevet français :

- 20 - N° 92 04322 du 03.04.1992, publiée sous le N° 2 689 519;  
- N° 92 13447 du 03.11.1992, publiée sous le N° 2 689 521;  
- N° 92 13443 du 03.11.1992, publiée sous le N° 2 689 520;  
- N° 94 01529 du 04.02.1994, publiée sous le N° 2 715 936;  
- N° 94 01531 du 04.02.1994, publiée sous le N° 2 715 939;  
25 - N° 94 01530 du 04.02.1994, publiée sous le N° 2 715 936;  
- N° 94 01532 du 04.02.1994, publiée sous le N° 2 715 937;  
- N° 94 14322 du 24.11.1994, publiée sous le N° 2 727 428;  
- N° 94 15810 du 23.12.1994, publiée sous le N° 2 728 585;  
et
- 30 - la demande de brevet WO-97/06260.

La présente invention concerne tout d'abord un matériel nucléaire, qui peut consister en un matériel rétroviral, à l'état isolé ou purifié, pouvant être appréhendé ou caractérisé de différentes manières :

- 35 - il comprend une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences

- SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117,  
SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130,  
SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii) les séquences  
complémentaires aux séquences (i) ; et (iii) les séquences  
5 équivalentes aux séquences (i) ou (ii), en particulier les  
séquences présentant pour toute suite de 100 monomères  
contigus, au moins 50 %, et préférentiellement au moins  
70 % d'homologie avec respectivement les séquences (i) ou  
(ii);
- 10 - il code pour un polypeptide présentant, pour  
toute suite contigue d'au moins 30 acides aminés, au  
moins 50 %, et de préférence au moins 70 % d'homologie,  
avec une séquence peptidique choisie dans le groupe qui  
consiste en SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118,  
15 SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137;
- son gène pol comprend une séquence nucléotidique  
identique ou équivalente à une séquence choisie dans le  
groupe qui consiste en SEQ ID NO: 112, SEQ ID NO: 124 et  
leurs séquences complémentaires;
- 20 - l'extrémité 5' de son gène pol commence au  
nucléotide 1419 de SEQ ID NO: 130;
- son gène pol code pour un polypeptide  
présentant, pour toute suite contigue d'au moins 30 acides  
aminés, au moins 50 %, et de préférence au moins 70 %  
25 d'homologie, avec la séquence peptidique SEQ ID NO: 113;
- l'extrémité 3' de son gène gag finit au  
nucléotide 1418 de SEQ ID NO: 130;
- son gène env comprend une séquence nucléotidique  
identique ou équivalente à une séquence choisie dans le  
30 groupe qui consiste en SEQ ID NO: 117, et ses séquences  
complémentaires;
- son gène env comprend une séquence nucléotidique  
qui commence au nucléotide 1 de SEQ ID NO: 117 et finit au  
nucléotide au nucléotide 233 de SEQ ID NO: 114;
- 35 - son gène env code pour un polypeptide  
présentant, pour toute suite contigue d'au moins 30 acides

aminées, au moins 50 %, et de préférence au moins 70 % d'homologie, avec la séquence SEQ ID NO: 118;

- la région U3R de son LTR 3' comprend une séquence nucléotidique qui se termine au nucléotide 617 de  
5 SEQ ID NO: 114;

- la région RU5 de son LTR 5' comprend une séquence nucléotidique qui commence au nucléotide 755 de SEQ ID NO: 120 et finit au nucléotide 337 de SEQ ID NO: 141 ou SEQ ID NO: 142;

10 - un matériel nucléique rétroviral comprenant une séquence qui commence au nucléotide 755 de SEQ ID NO: 120 et qui se termine au nucléotide 617 de SEQ ID NO: 114;

- le matériel nucléique rétroviral tel que défini précédemment est en particulier associé à au moins une  
15 maladie auto-immune telle que la sclérose en plaques ou la polyarthrite rhumatoïde.

L'invention concerne aussi un fragment nucléotidique qui répond au moins à l'une des définitions suivantes :

20 - il comprend ou consiste en une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii)  
25 les séquences complémentaires des séquences (i) ; et (iii) les séquences équivalentes aux séquences (i) ou (ii), en particulier les séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins 70 % d'homologie avec  
30 respectivement les séquences (i) ou (ii);

- il comprend ou consiste en une séquence nucléotidique codant pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie,  
35 avec une séquence peptidique choisie dans le groupe qui



consiste en SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137.

D'autres objets de la présente invention sont les suivants :

- 5           - une sonde nucléique pour la détection d'un rétrovirus associé à la sclérose en plaques et/ou la polyarthrite rhumatoïde, susceptible de s'hybrider spécifiquement sur tout fragment précédemment défini et appartenant au génome dudit rétrovirus; elle possède  
10           avantageusement de 10 à 100 nucléotides, de préférence de 10 à 30 nucléotides;
- une amorce pour l'amplification par polymérisation d'un ARN ou d'un ADN d'un rétrovirus associé à la sclérose en plaques et/ou la polyarthrite  
15           rhumatoïde, qui comprend une séquence nucléotidique identique ou équivalente à au moins une partie de la séquence nucléotidique d'un fragment défini précédemment, notamment une séquence nucléotidique présentant pour toute  
20           suite de 10 monomères contigus, au moins 50 %, de préférence au moins 70 % d'homologie avec au moins ladite partie dudit fragment ; de préférence la séquence nucléotidique d'une amorce de l'invention est choisie  
          parmi SEQ ID NO: 116, SEQ ID NO: 119, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 126, SEQ ID NO: 127,  
25           SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 132, et SEQ ID NO: 133;
- un ARN ou ADN, et notamment vecteur de réplication et/ou d'expression, comprenant un fragment génomique du matériel nucléique ou un fragment défini  
30           précédemment;
- un peptide codé par tout cadre de lecture ouvert appartenant à un fragment nucléotidique défini précédemment, notamment un polypeptide, par exemple  
          oligopeptide formant ou comprenant un déterminant  
35           antigénique reconnu par des sera de patients infectés par le virus MSRV-1, et/ou chez lesquels le virus MSRV-1 a été

réactivé; un peptide préférentiel comprend une séquence identique, partiellement ou totalement, ou équivalente à une séquence choisie parmi SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118; SEQ ID NO: 121, SEQ ID NO: 135 et  
5 SEQ ID NO: 137;

- une composition diagnostique, prophylactique, ou thérapeutique, notamment pour inhiber l'expression d'au moins un rétrovirus associé à la sclérose en plaques et/ou à la polyarthrite rhumatoïde, comprenant un fragment  
10 nucléotidique défini précédemment;

- un procédé pour détecter un rétrovirus associé à la sclérose en plaques et/ou à la polyarthrite rhumatoïde, dans un échantillon biologique, comprenant les étapes consistant à mettre en contact un ARN et/ou un ADN présumé  
15 appartenir ou provenant dudit rétrovirus, ou leur ARN et/ou ADN complémentaire, avec une composition comprenant un fragment nucléotidique défini ci-dessus.

Avant de détailler l'invention, différents termes utilisés dans la description et les revendications sont à  
20 présent définis :

- par souche ou isolat, on entend toute fraction biologique infectante et/ou pathogène, contenant par exemple des virus et/ou des bactéries et/ou des parasites, générant un pouvoir pathogène et/ou antigénique, hébergée  
25 par une culture ou un hôte vivant ; à titre d'exemple, une souche virale selon la définition précédente peut contenir un agent co-infectant, par exemple un protiste pathogène,

- le terme "MSRV" utilisé dans la présente description désigne tout agent pathogène et/ou infectant, associé à la SEP, notamment une espèce virale, les souches  
30 atténuées de ladite espèce virale, ou les particules défectives interférentes ou contenant des génomes co-encapsidés ou encore des génomes recombinés avec une partie du génome MSRV-1, dérivées de cette espèce. Il est  
35 connu que les virus et particulièrement les virus contenant de l'ARN ont une variabilité, consécutive

notamment à des taux relativement élevés de mutation spontanée, dont il sera tenu compte ci-après pour définir la notion d'équivalence,

- par virus humain, on entend un virus susceptible  
5 d'infecter ou d'être hébergé par l'être humain,

- compte tenu de toutes les variations et/ou recombinaison naturelles ou induites, pouvant être rencontrées dans la pratique de la présente invention, les objets de cette dernière, définis ci-dessus et dans les  
10 revendications, ont été exprimés en comprenant les équivalents ou dérivés des différents matériels biologiques définis ci-après, notamment des séquences homologues nucléotidiques ou peptidiques,

- le variant d'un virus ou d'un agent pathogène  
15 et/ou infectant selon l'invention, comprend au moins un antigène reconnu par au moins un anticorps dirigé contre au moins un antigène correspondant dudit virus et/ou dudit agent pathogène et/ou infectant, et/ou un génome dont toute partie est détectée par au moins une sonde  
20 d'hybridation, et/ou au moins une amorce d'amplification nucléotidique spécifique dudit virus et/ou agent pathogène et/ou infectant, dans des conditions d'hybridation déterminées bien connues de l'homme de l'art,

- selon l'invention, un fragment nucléotidique ou  
25 un oligonucléotide ou un polynucléotide est un enchaînement de monomères, ou un biopolymère, caractérisé par la séquence informationnelle des acides nucléiques naturels, susceptible de s'hybrider à tout autre fragment nucléotidique dans des conditions prédéterminées,  
30 l'enchaînement pouvant contenir des monomères de structures chimiques différentes et être obtenu à partir d'une molécule d'acide nucléique naturelle et/ou par recombinaison génétique et/ou par synthèse chimique ; un fragment nucléotidique peut être identique à un fragment  
35 génomique du virus MSRV-1 considéré par la présente

invention, notamment un gène de ce dernier, par exemple pol ou env dans le cas dudit virus ;

- ainsi un monomère peut être un nucléotide naturel d'acide nucléique, dont les éléments constitutifs sont un sucre, un groupement phosphate et une base azotée; dans l'ARN le sucre est le ribose, dans l'ADN le sucre est le désoxy-2-ribose; selon qu'il s'agit de l'ADN ou l'ARN, la base azotée est choisie parmi l'adénine, la guanine, l'uracile, la cytosine, la thymine; ou le nucléotide peut être modifié dans l'un au moins des trois éléments constitutifs ; à titre d'exemple, la modification peut intervenir au niveau des bases, générant des bases modifiées telles que l'inosine, la méthyl-5-désoxycytidine, la désoxyuridine, la diméthylaminéo-5-désoxyuridine, la diaminéo-2,6-purine, la bromo-5-désoxyuridine et toute autre base modifiée favorisant l'hybridation; au niveau du sucre, la modification peut consister dans le remplacement d'au moins un désoxyribose par un polyamide, et au niveau du groupement phosphate, la modification peut consister dans son remplacement par des esters, notamment choisis parmi les esters de diphosphate, d'alkyl et arylphosphonate et de phosphorothioate,

- par "séquence informationnelle", on entend toute suite ordonnée de monomères, dont la nature chimique et l'ordre dans un sens de référence, constituent ou non une information fonctionnelle de même qualité que celle des acides nucléiques naturels,

- par hybridation, on entend le processus au cours duquel, dans des conditions opératoires appropriées, deux fragments nucléotidiques, ayant des séquences suffisamment complémentaires, s'apparient pour former une structure complexe, notamment double ou triple, de préférence sous forme d'hélice,

- une sonde comprend un fragment nucléotidique synthétisé par voie chimique ou obtenu par digestion ou

coupure enzymatique d'un fragment nucléotidique plus long, comprenant au moins six monomères, avantageusement de 10 à 100 monomères, de préférence 10 à 30 monomères, et possédant une spécificité d'hybridation dans des conditions déterminées ; de préférence, une sonde possédant moins de 10 monomères n'est pas utilisée seule, mais l'est en présence d'autres sondes de taille aussi courte ou non ; dans certaines conditions particulières, il peut être utile d'utiliser des sondes de taille supérieure à 100 monomères ; une sonde peut notamment être utilisée à des fins de diagnostic et il s'agira par exemple de sondes de capture et/ou de détection,

- la sonde de capture peut être immobilisée sur un support solide par tout moyen approprié, c'est-à-dire directement ou indirectement, par exemple par covalence ou adsorption passive,

- la sonde de détection peut être marquée au moyen d'un marqueur choisi notamment parmi les isotopes radioactifs, des enzymes notamment choisis parmi la peroxydase et la phosphatase alcaline et ceux susceptibles d'hydrolyser un substrat chromogène, fluorigène ou luminescent, des composés chimiques chromophores, des composés chromogènes, fluorigènes ou luminescents, des analogues de bases nucléotidiques, et la biotine,

- les sondes utilisées à des fins de diagnostic de l'invention peuvent être mises en oeuvre dans toutes les techniques d'hybridation connues, et notamment les techniques dites "DOT-BLOT", "SOUTHERN BLOT", "NORTHERN BLOT" qui est une technique identique à la technique "SOUTHERN BLOT" mais qui utilise de l'ARN comme cible, la technique SANDWICH ; avantageusement, on utilise la technique SANDWICH dans la présente invention, comprenant une sonde de capture spécifique et/ou une sonde de détection spécifique, étant entendu que la sonde de capture et la sonde de détection doivent présenter une séquence nucléotidique au moins partiellement différente,

- toute sonde selon la présente invention peut s'hybrider in vivo ou in vitro sur l'ARN et/ou sur l'ADN, pour bloquer les phénomènes de répllication, notamment traduction et/ou transcription, et/ou pour dégrader ledit  
5 ADN et/ou ARN,

- une amorce est une sonde comprenant au moins six monomères, et avantageusement de 10 à 30 monomères, possédant une spécificité d'hybridation dans des conditions déterminées, pour l'initiation d'une  
10 polymérisation enzymatique, par exemple dans une technique d'amplification telle que la PCR (Polymerase Chain Reaction), dans un procédé d'élongation, tel que le séquençage, dans une méthode de transcription inverse ou analogue,

15 - deux séquences nucléotidiques ou peptidiques sont dites équivalentes ou dérivées l'une par rapport à l'autre, ou par rapport à une séquence de référence, si fonctionnellement les biopolymères correspondants peuvent jouer sensiblement le même rôle, sans être identiques,  
20 vis-à-vis de l'application ou utilisation considérée, ou dans la technique dans laquelle elles interviennent ; sont notamment équivalentes deux séquences obtenues du fait de la variabilité naturelle, notamment mutation spontanée de l'espèce à partir de laquelle elles ont été identifiées,  
25 ou induite, ainsi que deux séquences homologues, l'homologie étant définie ci-après,

- par "variabilité", on entend toute modification, spontanée ou induite d'une séquence, notamment par substitution, et/ou insertion, et/ou délétion de  
30 nucléotides et/ou de fragments nucléotidiques, et/ou extension et/ou raccourcissement de la séquence à l'une au moins des extrémités; une variabilité non naturelle peut résulter des techniques de génie génétique utilisées, par exemple du choix des amorces de synthèse dégénérées ou  
35 non, retenues pour amplifier un acide nucléique; cette variabilité peut se traduire par des modifications de

toute séquence de départ, considérée comme référence, et pouvant être exprimées par un degré d'homologie par rapport à ladite séquence de référence,

- l'homologie caractérise le degré d'identité de  
5 deux fragments nucléotidiques ou peptidiques comparés ; elle se mesure par le pourcentage d'identité qui est notamment déterminé par comparaison directe de séquences nucléotidiques ou peptidiques, par rapport à des séquences nucléotidiques ou peptidiques de référence,
- 10 - tout fragment nucléotidique est dit équivalent ou dérivé d'un fragment de référence, s'il présente une séquence nucléotidique équivalente à la séquence du fragment de référence ; selon la définition précédente, sont notamment équivalents à un fragment nucléotidique de  
15 référence :
  - (a) tout fragment susceptible de s'hybrider au moins partiellement avec le complément du fragment de référence,
  - (b) tout fragment dont l'alignement avec le  
20 fragment de référence conduit à mettre en évidence des bases contigues identiques, en nombre plus important qu'avec tout autre fragment provenant d'un autre groupe taxonomique,
  - (c) tout fragment résultant ou pouvant résulter de  
25 la variabilité naturelle de l'espèce, à partir de laquelle il est obtenu,
  - (d) tout fragment pouvant résulter des techniques de génie génétique appliquées au fragment de référence,
  - (e) tout fragment, comportant au moins huit  
30 nucléotides contigus, codant pour un peptide homologue ou identique au peptide codé par le fragment de référence,
  - (f) tout fragment différent du fragment de référence, par insertion, déletion, substitution d'au moins un monomère, extension, ou raccourcissement à l'une  
35 au moins de ses extrémités ; par exemple, tout fragment correspondant au fragment de référence, flanqué à l'une au

moins de ses extrémités par une séquence nucléotidique ne codant pas pour un polypeptide,

- par polypeptide, on entend notamment tout peptide d'au moins deux acides aminés, notamment  
5 oligopeptide, protéine, extrait, séparé, ou substantiellement isolé ou synthétisé, par l'intervention de la main de l'homme, notamment ceux obtenus par synthèse chimique, ou par expression dans un organisme recombinant,

- par polypeptide codé de manière partielle par un  
10 fragment nucléotidique, on entend un polypeptide présentant au moins trois acides aminés codés par au moins neuf monomères contigus compris dans ledit fragment nucléotidique,

- un acide aminé est dit analogue à un autre  
15 acide aminé, lorsque leurs caractéristiques physico-chimiques respectives, telles que polarité, hydrophobicité, et/ou basicité, et/ou acidité, et/ou neutralité, sont sensiblement les mêmes ; ainsi, une leucine est analogue à une isoleucine.

20 - tout polypeptide est dit équivalent ou dérivé d'un polypeptide de référence, si les polypeptides comparés ont sensiblement les mêmes propriétés, et notamment les mêmes propriétés antigéniques, immunologiques, enzymologiques et/ou de reconnaissance  
25 moléculaire ; est notamment équivalent à un polypeptide de référence :

(a) tout polypeptide possédant une séquence dont au moins un acide aminé a été substitué par un acide aminé analogue,

30 (b) tout polypeptide ayant une séquence peptidique équivalente, obtenue par variation naturelle ou induite dudit polypeptide de référence, et/ou du fragment nucléotidique codant pour ledit polypeptide,

(c) un mimotope dudit polypeptide de référence,



(d) tout polypeptide dans la séquence duquel un ou plusieurs acides aminés de la série L sont remplacés par un acide aminé de la série D, et vice versa,

(e) tout polypeptide dans la séquence duquel on a  
5 introduit une modification des chaînes latérales des acides aminés, telle que par exemple une acétylation des fonctions aminées, une carboxylation des fonctions thiol, une estérification des fonctions carboxyliques,

(f) tout polypeptide dans la séquence duquel une  
10 ou des liaisons peptidiques ont été modifiées, comme par exemple les liaisons carba, rétro, inverso, rétro-inverso, réduites, et méthylène-oxy,

(g) tout polypeptide dont au moins un antigène est  
reconnu par anticorps dirigé contre un polypeptide de  
15 référence,

- le pourcentage d'identité caractérisant l'homologie de deux fragments peptidiques comparés est selon la présente invention d'au moins 50% et de préférence au moins 70 %.

20 Etant donné qu'un virus possédant une activité enzymatique transcriptase inverse peut être génétiquement caractérisé aussi bien sous forme d'ARN que d'ADN, il sera fait mention aussi bien de l'ADN que de l'ARN viral pour caractériser les séquences relatives à un virus possédant  
25 une telle activité transcriptase inverse, dit MSRV-1 selon la présente description.

Les expressions d'ordre utilisées dans la présente description et les revendications, telles que "première séquence nucléotidique" ne sont pas retenues pour exprimer  
30 un ordre particulier, mais pour définir plus clairement l'invention.

Par détection d'une substance ou agent, on entend ci-après aussi bien une identification, qu'une quantification, ou une séparation ou isolement de ladite  
35 substance ou dudit agent.

L'invention sera mieux comprise à la lecture de la description détaillée qui va suivre faite en référence aux figures annexées dans lesquelles :

La Figure 1 représente la structure générale de l'ADN proviral et l'ARN génomique de MSRV-1.

La Figure 2 représente la séquence nucléotidique du clone dénommé CL6-5' (SEQ ID NO: 112) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 3 représente la séquence nucléotidique du clone dénommé CL6-3' (SEQ ID NO: 114) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 4 représente la séquence nucléotidique du clone dénommé C15 (SEQ ID NO: 117) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 5 représente la séquence nucléotidique du clone dénommé 5M6 (SEQ ID NO: 120) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 6 représente la séquence nucléotidique du clone dénommé CL2 (SEQ ID NO: 130) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 7 représente trois trames de lecture potentielles en acides aminés exprimées par pET28C-clone 2 et figurant sous la séquence nucléotidique.

La Figure 8 représente trois trames de lecture potentielles en acides aminés exprimées par pET21C-clone 2 et figurant sous la séquence nucléotidique.

La Figure 9 représente la séquence nucléotidique du clone dénommé LB13 (SEQ ID NO: 141) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 10 représente la séquence nucléotidique du clone dénommé LA15 (SEQ ID NO: 142) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

5 La Figure 11 représente la séquence nucléotidique du clone dénommé LB16 (SEQ ID NO: 124) et trois trames de lecture potentielles en acides aminés figurant sous la séquence nucléotidique.

La Figure 12 représente l'activité promotrice  
10 exprimée en cpm/4min des séquences U3R sous clonées à partir de LTRs d'origines différentes dans le plasmide PCAT3. PCAT3 signifie plasmide seul, PCAT-PH74 signifie plasmide plus clone U3Rendogène exprimé dans le placenta, PCAT-cl6 signifie plasmide plus clone U3R amplifié dans  
15 l'ARN d'un plasma SEP, PCAT-5M6 signifie plasmide plus région U3R amplifiée dans l'ADN cellulaire, "no plasmid" signifie absence de plasmide dans le test.

La Figure 13 représente les séquences MSRV1 env et LTR3'. Les flèches horizontales indiquent le début des  
20 régions env, U3 et R. Dans la région env : le peptide signal et la région immunosuppressive potentielle sont soulignés, les sites de glycosilation potentiels sont encadrés et les sites de clivage potentiels sont indiqués par des flèches verticales. Dans la région U3R : l'élément  
25 de régulation CAAT et la TATA Box sont soulignés, le site "cap" et le signal de polyadénylation sont aussi indiqués.

La Figure 14 représente une région LTR5' (RU5) suivie d'un site PBS (primer binding site) complémentaire  
du tARN Trp et d'un gène gag codant pour une protéine  
30 d'environ 487 acides aminés. Les acides aminés conservés dans la nucléocapside sont soulignés deux fois. Les acides aminés définissant la région de plus forte homologie dans la capside sont en gras et soulignés une fois. Les / dans la séquence en acides aminés indiquent des variations  
35 observées selon les clones et dans la séquence nucléotidique ils indiquent des sauts de trame dans

certaines clones. Les régions encadrées correspondent à des épitopes identifiés par analyse peptidique de la région C-terminale.

La Figure 15 représente la région intégrase de  
5 MSRV1, la séquence nucléotidique et la séquence en acides aminés déduite de la région intégrase correspondant au clone 87-23. Dans la Figure 15 // signifie un saut de trame qui a été supprimé pour restituer l'ORF potentiel. Les lettres en caractères gras soulignées représentent les  
10 acides aminés conservés des intégrases rétrovirales.

La Figure 16 décrit les séquences nucléotidique et peptidique du clone B13 (identique au clone FBd13 décrit dans des demandes antérieures) avec indication des ORFs et des codons stop représentés par un point. La région  
15 soulignée en gras représente le domaine immunosuppresseur potentiel. Le domaine souligné en simple représente le début du LTR3'.

**EXEMPLE 1:** OBTENTION D'UNE REGION CL6-5' CODANT POUR  
20 L'EXTREMITÉ N-TERMINALE DE L'INTEGRASE ET D'UNE REGION CL6-3' CONTENANT LA SEQUENCE 3' TERMINALE DU GENOME MSRV-1

Une 3'RACE a été effectuée sur de l'ARN total extrait de plasma d'un patient atteint de SEP. Un plasma  
25 témoin sain, traité sous les mêmes conditions, a été utilisé comme contrôle négatif. La synthèse de cDNA a été réalisée avec une amorce oligo dT identifiée par SEQ ID NO: 68 (5' GAC TCG CTG CAG ATC GAT TTT TTT TTT TTT TTT To 3') et la transcriptase inverse "Expand<sup>TM</sup> RT" de  
30 Boehringer selon les conditions préconisées par la société. Une PCR a été effectuée avec l'enzyme Klentaq (Clontech) sous les conditions suivantes : 94°C 5 min puis 93°C 1 min, 58°C 1 min, 68°C 3 min pendant 40 cycles et 68°C pendant 8 min, avec un volume réactionnel final de  
35 50 µl.

Amorces utilisées pour la PCR:

- amorce 5', identifiée par SEQ ID NO: 69  
5' GCC ATC AAG CCA CCC AAG AAC TCT TAA CTT 3' ;
- amorce 3', identifiée par SEQ ID NO: 68

Une deuxième PCR dite "semi-nichée" a été  
5 réalisée avec une amorce 5' située à l'intérieur de la  
région déjà amplifiée. Cette deuxième PCR a été effectuée  
sous les mêmes conditions expérimentales que celles  
utilisées lors de la première PCR, en utilisant 10  $\mu$ l du  
produit d'amplification issu de la première PCR.

10 Amorces utilisées pour la PCR semi-nichée:

- amorce 5', identifiée par SEQ ID NO: 70  
5' CCA ATA GCC AGA CCA TTA TAT ACA CTA ATT 3' ;
- amorce 3', identifiée par SEQ ID NO: 68

Les amorces SEQ ID NO: 69 et SEQ ID NO: 70 sont  
15 spécifiques de la région pol de MSRV-1.

Un produit d'amplification de 1,9Kb a été obtenu  
pour le plasma de patient SEP. Le fragment correspondant  
n'a pas été observé pour le plasma témoin sain. Ce produit  
d'amplification a été cloné de la façon suivante :

20 l'ADN amplifié a été inséré dans un plasmide à l'aide du  
kit TA Cloning®. Les 2  $\mu$ l de solution d'ADN ont été  
mélangés avec 5  $\mu$ l d'eau distillée stérile, 1  $\mu$ l d'un  
tampon de ligation 10 fois concentré "10X LIGATION  
25 BUFFER", 2  $\mu$ l de "PCR<sup>TM</sup> VECTOR" (25 ng/ml), et 1  $\mu$ l de "T4  
DNA LIGASE". Ce mélange a été incubé une nuit à 12°C. Les  
étapes suivantes ont été réalisées conformément aux  
instructions du kit TA Cloning® (Invitrogen). A la fin de  
la procédure, les colonies blanches de bactéries  
recombinantes (white) ont été repiquées pour être  
30 cultivées et permettre l'extraction des plasmides  
incorporés selon la procédure dite de "miniprep". La  
préparation de plasmide de chaque colonie recombinante a  
été coupée par une enzyme de restriction appropriée et  
analysée sur gel d'agarose. Les plasmides possédant un  
35 insert détecté sous lumière UV après marquage du gel au  
bromure d'éthidium, ont été sélectionnés pour le

séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur Sp6 présent sur le plasmide de clonage du TA cloning kit®. La réaction préalable au séquençage a ensuite été effectuée selon la méthode préconisée pour l'utilisation du kit de séquençage "PRISM™ Ready Reaction AmpliTaq® FS, DyeDeoxy™ Terminator" (Applied Biosystems, réf. 402119) et le séquençage automatique a été réalisé sur les appareils 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

Le clone obtenu, contient une région CL6-5'codant pour l'extrémité N terminale de l'intégrase et une région CL6-3', correspondant à la région 3' terminale de MSRV-1 et permettant de définir la fin de l'enveloppe (234 pb) et les régions U3, R (401 pb) du rétrovirus MSRV1.

La région correspondant à l'extrémité N terminale de l'intégrase est représentée par sa séquence nucléotidique (SEQ ID NO: 112) dans la figure 27. Les trois trames de lecture potentielles sont présentées par leur séquence aminéoacide sous la séquence nucléotidique, et la séquence aminéoacide de l'extrémité N-terminale de l'intégrase est identifiée par SEQID NO: 113.

La région Cl6-3' est représentée par sa séquence nucléotidique (SEQ ID NO: 114) dans la figure 3. Les trois trames de lecture potentielles sont présentées par leur séquence aminéoacide sous la séquence nucléotidique. Une séquence aminéoacide correspondant à l'extrémité C-terminale de la protéine env de MSRV-1 est identifiée par SEQ ID NO: 115.

Afin d'évaluer l'activité promotrice du LTR obtenu à partir de clone 6 (cl6) un test d'activité promotrice utilisant l'enzyme CAT (chloramphénicol acetyl transferase) a été effectué avec la région U3R correspondante. En parallèle un clone contenant la même région U3R d'ARN rétroviral endogène exprimé dans le placenta normal (PH74) et un clone (5M6) provenant d'ADN

ont été testés. Le résultat présenté dans la figure 12 montre une très forte activité promotrice du LTR issu de plasma SEP (c16) et une activité significativement beaucoup plus faible avec les séquences d'origine endogène non SEP.

**EXEMPLE 2: OBTENTION DU CLONE C15 CONTENANT LA REGION CODANT POUR UNE PARTIE DE L'ENVELOPPE DU RETROVIRUS MSRV-1**

Une RT-PCR a été effectuée sur de l'ARN total extrait de virions concentrés par ultracentrifugation à partir d'un surnageant de culture de synoviocytes provenant d'un patient PR. La synthèse de cDNA a été réalisée avec une amorce oligo dT et la transcriptase inverse "Expand<sup>TM</sup> RT" de Boehringer selon les conditions préconisées par la société. Une PCR a été effectuée avec l'Expand<sup>TM</sup> Long Template PCR System (Boehringer) sous les conditions suivantes : 94°C 5 min puis 93°C 1 min, 60°C 1 min, 68°C 3 min pendant 40 cycles et 68°C pendant 8 min et avec un volume réactionnel final de 50 µl.

Amorces utilisées pour la PCR:

- amorce 5', identifiée par SEQ ID NO: 69

5' GCC ATC AAG CCA CCC AAG AAC TCT TAA CTT 3' ;

- amorce 3', identifiée par SEQ ID NO: 116

5' TGG GGT TCC ATT TGT AAG ACC ATC TGT AGC TT 3'

Une deuxième PCR dite "semi-nichée" a été réalisée avec une amorce 5' située à l'intérieur de la région déjà amplifiée. Cette deuxième PCR a été effectuée sous les mêmes conditions expérimentales que celles utilisées lors de la première PCR (sauf que 30 cycles ont été réalisés au lieu de 40), en utilisant 10 µl du produit d'amplification issu de la première PCR.

Amorces utilisées pour la PCR semi-nichée:

- amorce 5', identifiée par SEQ ID NO: 70

5' CCA ATA GCC AGA CCA TTA TAT ACA CTA ATT 3' ;

- amorce 3', identifiée par SEQ ID NO: 116

Les amorces SEQ ID NO: 69 et SEQ ID NO: 70 sont spécifiques de la région pol de MSRV-1. L'amorce SEQ ID NO: 116 est spécifique de la séquence FBd13 (aussi dénommé B13) et est localisée dans la région env conservée parmi les oncorétrovirus.

Un produit d'amplification de 1932 pb a été obtenu et cloné de la façon suivante :  
l'ADN amplifié a été inséré dans un plasmide à l'aide du kit TA Cloning®. Les différentes étapes ont été réalisées conformément aux instructions du kit TA Cloning® (Invitrogen). A la fin de la procédure, les colonies blanches de bactéries recombinantes (white) ont été repiquées pour être cultivées et permettre l'extraction des plasmides incorporés selon la procédure dite de "miniprep". La préparation de plasmide de chaque colonie recombinante a été coupée par une enzyme de restriction appropriée et analysée sur gel d'agarose. Les plasmides possédant un insert détecté sous lumière UV après marquage du gel au bromure d'éthidium, ont été sélectionnés pour le séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur SP6 présent sur le plasmide de clonage du TA cloning kit®. La réaction préalable au séquençage a ensuite été effectuée selon la méthode préconisée pour l'utilisation du kit de séquençage "PRISM™ Ready Reaction AmpliTaq® FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les appareils 373A et 377 Applied Biosystems, selon les instructions du fabricant.

Le clone C15 obtenu, contient une région correspondant à la région de l'enveloppe de MSRV-1, de 1481 pb.

La région env du clone C15 est représentée par sa séquence nucléotidique (SEQ ID NO: 117) dans la figure 5. Les trois trames de lecture potentielles de ce clone sont présentées par leur séquence aminéoacide sous la séquence



nucléotidique. La trame de lecture correspondant à une protéine env structurale MSRV-1 est identifiée par SEQ ID NO: 118.

A partir des séquences définies provenant des clones cl6 et C15, une construction plasmidique a pu être effectuée codant pour une enveloppe complète suivie du LTR3', comme présenté dans la figure 13 avec la trame de lecture correspondante.

10 **EXEMPLE 3:** OBTENTION D'UN CLONE 5M6 CONTENANT LES SEQUENCES DE LA REGION 3' TERMINALE DE L'ENVELOPPE, SUIVIES DES SEQUENCES U3,R,U5 DE TYPE PROVIRAL MSRV-1.

Une PCR monodirectionnelle a été effectuée sur de l'ADN extrait de lymphocytes B immortalisés en culture d'un patient PR. La PCR a été effectuée avec l'Expand™ Long Template PCR System (Boehringer) sous les conditions suivantes : 94°C 3 min puis 93°C 1 min, 60°C 1 min, 68°C 3 min pendant 10 cycles , puis 93°C 1 min, 60°C 1 min avec 15 sec d'extension à chaque cycle, 68°C 3 min pendant 35 cycles et 68°C pendant 7 min et avec un volume réactionnel final de 50 µl.

L'amorce utilisée pour la PCR identifiée par SEQ ID NO: 119 est 5' TCA AAA TCG AAG AGC TTT AGA CTT GCT AAC CG 3' ;

L'amorces SEQ ID NO: 119 est spécifique de la région env du clone C15.

Un produit d'amplification de 1673 pb a été obtenu et cloné de la façon suivante :

l'ADN amplifié a été inséré dans un plasmide à l'aide du kit TA Cloning®. Les différentes étapes ont été réalisées conformément aux instructions du kit TA Cloning® (Invitrogen). A la fin de la procédure, les colonies blanches de bactéries recombinantes (white) ont été repiquées pour être cultivées et permettre l'extraction des plasmides incorporés selon la procédure dite de

"miniprep". La préparation de plasmide de chaque colonie recombinante a été coupée par une enzyme de restriction appropriée et analysée sur gel d'agarose. Les plasmides possédant un insert détecté sous lumière UV après marquage  
5 du gel au bromure d'éthidium, ont été sélectionnés pour le séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur T7 présent sur le plasmide de clonage du TA cloning kit®. La réaction préalable au séquençage a ensuite été effectuée selon la méthode  
10 préconisée pour l'utilisation du kit de séquençage "PRISM™ Ready Reaction AmpliTaq® FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les appareils 373 A et 377 Applied Biosystems, selon les instructions du  
15 fabricant.

Le clone 5M6 obtenu, contient une région correspondant à la région 3' de l'enveloppe de MSRV-1, de 492 pb suivi des régions U3, R et U5 (837 pb) de MSRV1.

Le clone 5M6 est représenté par sa séquence  
20 nucléotidique (SEQ ID NO: 120) dans la figure 5. Les trois trames de lecture potentielles de ce clone sont présentées par leur séquence aminéoacide sous la séquence nucléotidique. La trame de lecture correspondant à l'extrémité C-terminale de la protéine env MSRV-1 est  
25 identifiée par SEQ ID NO: 121.

**EXEMPLE 4: OBTENTION DU CLONE LB16 CONTENANT LA REGION CODANT L'INTEGRASE DU RETROVIRUS MSRV-1**

Une RT-PCR a été effectuée sur de l'ARN total  
30 traité à la DNaseI et extrait à partir d'un plexus choroïde provenant d'un patient SEP. La synthèse de cDNA a été réalisée avec une amorce oligo dT et la transcriptase inverse "Expand™ RT" de Boehringer selon les conditions préconisées par la société. Un contrôle "no RT" a été  
35 effectué parallèlement sur le même matériel. Une PCR a été effectuée avec la Taq polymérase (Perkin Elmer) sous les

conditions suivantes : 95°C 5 min puis 95°C 1 min, 55°C 1 min, 72°C 2 min pendant 35 cycles et 72°C pendant 8 min et avec un volume réactionnel final de 50 µl.

Amorces utilisées pour la PCR:

5 - amorce 5', identifiée par SEQ ID NO: 122

5' GGC ATT GAT AGC ACC CAT CAG 3' ;

- amorce 3', identifiée par SEQ ID NO: 123

5' CAT GTC ACC AGG GTG GAA TAG 3'

L'amorce SEQ ID NO: 122 est spécifique de la région pol de MSRV-1 et plus précisément similaire à la région intégrase décrite précédemment. L'amorce SEQ ID NO 123 a été définie sur des séquences des clones obtenus lors d'essais préalables.

Un produit d'amplification d'environ 760 pb a été obtenu uniquement dans l'essai avec RT et a été cloné de la façon suivante :

L'ADN amplifié a été inséré dans un plasmide à l'aide du kit TA Cloning®. Les différentes étapes ont été réalisées conformément aux instructions du kit TA Cloning® (Invitrogen). A la fin de la procédure, les colonies blanches de bactéries recombinantes (white) ont été repiquées pour être cultivées et permettre l'extraction des plasmides incorporés selon la procédure dite de "miniprep". La préparation de plasmide de chaque colonie recombinante a été coupée par une enzyme de restriction appropriée et analysée sur gel d'agarose. Les plasmides possédant un insert détecté sous lumière UV après marquage du gel au bromure d'éthidium, ont été sélectionnés pour le séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur T7 présent sur le plasmide de clonage du TA cloning kit®. La réaction préalable au séquençage a ensuite été effectuée selon la méthode préconisée pour l'utilisation du kit de séquençage "PRISM™ Ready Reaction AmpliTaq® FS, DyeDeoxy™ Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les appareils

373 A et 377 Applied Biosystems, selon les instructions du fabricant.

Le clone LB16 obtenu, contient les séquences correspondant à l'intégrase. La séquence nucléotidique de ce clone est identifiée par SEQ ID NO: 124 sur la figure 11, trois trames de lecture sont déterminées.

**EXEMPLE 5:** OBTENTION D'UN CLONE 2, CL2, CONTENANT EN 3' UNE PARTIE HOMOLOGUE AU GENE POL, CORRESPONDANT AU GENE PROTEASE, ET AU GENE GAG (GM3) CORRESPONDANT A LA NUCLEOCAPSIDE, ET UNE NOUVELLE REGION 5' CODANTE, CORRESPONDANT AU GENE GAG PLUS SPECIFIQUEMENT LA MATRICE ET LA CAPSIDE de MSRV-1.

15 Une amplification par PCR a été effectuée sur de l'ARN total extrait à partir de 100 µl de plasma d'un patient atteint de SEP. Un témoin eau, traité sous les mêmes conditions, a été utilisé comme contrôle négatif. La synthèse de cDNA a été réalisée avec 300 pmole d'une  
20 amorce aléatoire (GIBCO-BRL, France) et la transcriptase inverse "Expand RT" (BOEHRINGER MANNHEIM, France) selon les conditions préconisées par la société. Une amplification par PCR ("polymerase chain reaction") a été effectuée avec l'enzyme Taq polymérase (Perkin Elmer,  
25 France) en utilisant 10 µl de cDNA sous les conditions suivantes: 94°C 2 min, 55°C 1 min, 72°C 2 min puis 94°C 1 min, 55°C 1 min, 72°C 2 min pendant 30 cycles et 72°C pendant 7 min et avec un volume réactionnel final de 50 µl.

30 Amorces utilisées pour l'amplification par PCR:

- amorce 5', identifiée par SEQ ID NO: 126

5' CGG ACA TCC AAA GTG ATG GGA AAC G 3' ;

- amorce 3', identifiée par SEQ ID NO: 127

5' GGA CAG GAA AGT AAG ACT GAG AAG GC 3'

35 Une deuxième amplification par PCR dite "semi-nichée" a été réalisée avec une amorce 5' située à

l'intérieur de la région déjà amplifiée. Cette deuxième PCR a été effectuée sous les mêmes conditions expérimentales que celles utilisées lors de la première PCR, en utilisant 10  $\mu$ l du produit d'amplification issu de la première PCR.

Amorces utilisées pour l'amplification par PCR semi-nichée:

- amorce 5', identifiée par SEQ ID NO: 128  
5' CCT AGA ACG TAT TCT GGA GAA TTG GG 3' ;
- 10 - amorce 3', identifiée par SEQ ID NO: 129  
5' TGG CTC TCA ATG GTC AAA CAT ACC CG 3'

Les amorces SEQ ID NO: et SEQ ID NO: sont spécifiques de la région pol, clone G+E+A, plus spécifiquement la région E: position nucléotidique n° 423 à n° 448. Les amorces utilisées dans la région 5' ont été définies sur des séquences de clones obtenus lors d'essais préalables.

Un produit d'amplification de 1511 pb a été obtenu à partir de l'ARN extrait du plasma de patient SEP. Le fragment correspondant n'a pas été observé pour le témoin eau. Ce produit d'amplification a été cloné de la façon suivante.

L'ADN amplifié a été inséré dans un plasmide à l'aide du kit TA Cloning<sup>TM</sup>. Les 2  $\mu$ l de solution d'ADN ont été mélangés avec 5  $\mu$ l d'eau distillée stérile, 1  $\mu$ l d'un tampon de ligation 10 fois concentré "10X LIGATION BUFFER", 2  $\mu$ l de "PCR<sup>TM</sup> VECTOR" (25 ng/ml) et 1  $\mu$ l de "T4 DNA LIGASE". Ce mélange a été incubé une nuit à 14°C. Les étapes suivantes ont été réalisées conformément aux instructions du kit TA Cloning® (Invitrogen). Le mélange a été étalé après transformation de la ligation dans des bactéries *E. coli* INVαF'. A la fin de la procédure, les colonies blanches de bactéries recombinantes ont été repiquées pour être cultivées et permettre l'extraction des plasmides incorporés selon la procédure dite de "minipréparation d'ADN" (17). La préparation de plasmide

de chaque colonie recombinante a été coupée par l'enzyme de restriction Eco RI et analysée sur gel d'agarose. Les plasmides possédant un insert détecté sous lumière UV après marquage du gel au bromure d'éthidium, ont été  
5 sélectionnés pour le séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur T7 présent sur le plasmide de clonage du TA cloning kit®. La réaction préalable au séquençage a ensuite été effectuée selon la méthode préconisée pour l'utilisation du kit de  
10 séquençage "PRISM<sup>TM</sup> Ready Reaction Amplitaq® FS, DyeDeoxy<sup>TM</sup> Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les appareils 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

15 Le clone obtenu, dénommé CL2, contient une région C-terminale similaire à la région 5' terminale des clones G+E+A de MSRV-1, qui permet de définir la région C-terminale du gène gag et une nouvelle région correspondante à la région N-terminale du gène gag MSRV-1.

20 CL2 permet de définir une région de 1511 pb présentant une phase ouverte de lecture dans la région N-terminale de 1077 pb codante pour 359 acides aminés et une phase non-ouverte de lecture, de 454 pb, correspondant à la région C-terminale du gène gag MSRV-1.

25 La séquence nucléotidique de CL2 est identifiée par SEQ ID NO: 130. Elle est représentée à la figure 6, avec les trames de lecture potentielles en aminéoacide.

Le fragment de 1077 pb de CL2 codant pour 359 acides aminés a été amplifié par PCR avec l'enzyme Pwo  
30 (5U/ $\mu$ l) (Boehringer-Mannheim, France) en utilisant 1  $\mu$ l de la minipréparation de l'ADN du clone 2 sous les conditions suivantes: 95°C 1 min, 60°C 1 min, 72°C 2 min pendant 25 cycles et avec un volume réactionnel final de 50  $\mu$ l à l'aide des amorcés:

35 - amorce 5' (Bam HI), identifié par SEQ ID NO: 132

5' TGC TGG AAT TCG GGA TCC TAG AAC GTA TTC 3' (30 mer), et  
 - amorce 3' (*Hind* III), identifié par SEQ ID NO: 133  
 5 AGT TCT GCT CCG AAG CTT AGG CAG ACT TTT 3' (30 mer)  
 correspondant, respectivement, à la séquence nucléotidique  
 du clone 2 en position -9 à 21 et 1066 à 1095.

Le fragment obtenu après PCR, a été linéarisé par  
*Bam* HI et *Hind* III, et sous-cloné dans les vecteurs  
 d'expression pET28C et pET21C (NOVAGEN) linéarisé par *Bam*  
 HI et *Hind* III. Le séquençage de l'ADN du fragment de 1077  
 pb du clone 2 dans les deux vecteurs d'expression a été  
 réalisé selon la méthode préconisée pour l'utilisation du  
 kit de séquençage "PRISM<sup>TM</sup> Ready Reaction Amplitaq<sup>®</sup> FS,  
 DyeDeoxy<sup>TM</sup> Terminator" (Applied Biosystems, ref. 402119)  
 et le séquençage automatique a été réalisé sur les  
 appareils 373 A et 377 Applied Biosystems, selon les  
 instructions du fabricant.

L'expression de la séquence nucléotidique du  
 fragment de 1077 pb du clone 2 par les vecteurs  
 d'expression pET28C et pET21C sont identifiées par  
 respectivement SEQ ID NO: 135 et SEQ ID NO: 137.

#### **EXEMPLE 6: L'EXPRESSION DU CLONE 2 CHEZ *ESCHERICHIA COLI***

Les constructions pET28c-clone-2 (1077 pb) et  
 pET21C-clone 2 (1077 pb) synthétisent, dans la souche  
 bactérienne BL21 (DE3), une protéine en fusion N- et C-  
 terminale pour le vecteur pET28C et C-Terminalé pour le  
 vecteur pET21C avec 6 Histidines; de masse moléculaire  
 apparente d'environ 45 kDa, mise en évidence par  
 électrophorèse sur gel de polyacrylamide (SDS-PAGE) (SDS =  
 Dodecyl Sulfate de Sodium) (Laemmli, 1970, (1)). La  
 réactivité de la protéine a été mise en évidence vis à vis  
 d'un anticorps monoclonal anti-Histidine (DIANOVA) par la  
 technique de Western blot (Towbin et al., 1979, (2)).

Les protéines recombinantes pET28c-clone 2  
 (1077 pb) et pET21C-clone 2 (1077 pb) ont été visualisées

en SDS-PAGE dans la fraction insoluble après digestion enzymatique des extraits bactériens avec 50  $\mu$ l de lysozyme (10 mg/ml) et lyse par ultrasons.

Les propriétés antigéniques des antigènes recombinants pET28C-clone 2 (1077 pb) et pET21C-clone 2 (1077 pb) ont été testées par Western Blot () après solubilisation du culot bactérien avec 2% SDS et 50 mM  $\beta$ -mercaptoéthanol. Après incubation avec les sérums de patients atteints de sclérose en plaques, les sérums des témoins neurologiques et les sérums de témoins de centre de transfusion sanguine (CTS), les immunocomplexes ont été détectés à l'aide d'un sérum de chèvre anti-IgG et anti IgM humaines, couplé à la phosphatase alcaline.

Les résultats sont présentés dans le tableau ci-après.

TABLEAU

Réactivité de sérums atteints de sclérose en plaques et témoins avec la protéine recombinante MSRV-1 gag clone 2 (1077 pb) = pET21C-clone 2 (1077 pb) et pET28C-clone 2 (1077 pb)<sup>a</sup>

| MALADIE           | NOMBRE<br>D'INDIVIDUS TESTÉS | NOMBRE<br>D'INDIVIDUS POSITIFS |
|-------------------|------------------------------|--------------------------------|
|                   |                              |                                |
| SEP               | 15                           | 6                              |
| 25<br>TEMOINS     |                              | 2(+++), 2(++), 2(+)            |
| NEUROLOGIQUES     | 2                            | 1(+++)                         |
| TEMOINS           |                              |                                |
| 30<br>SAINS (CTS) | 22                           | 1(+/-)                         |

(a) Les bandelettes contenant 1,5  $\mu$ g d'antigène recombinant pET-gag clone 2 (1077 pb) présentent une réactivité contre de sérums dilués au 1/100. L'interprétation de Western Blot est basée sur la présence ou absence d'une bande pET-gag clone 2 (1077 pb)



spécifique sur les bandelettes. Des contrôles positifs et négatifs sont inclus dans chaque expérience.

Ces résultats montrent que, dans les conditions techniques utilisées, environ 40% des sérums humains atteints de sclérose en plaques testés réagissent avec les protéines recombinantes pET28C-clone 2 (1077 pb) et pET21C-clone 2 (1077 pb). Une réactivité a été observée sur un témoin neurologique et il est intéressant de noter que les ARN extraits à partir de ce sérum, après l'étape de transcriptase inverse, sont aussi amplifiés par PCR dans la région pol. Ceci suggère que des personnes n'ayant pas déclaré une SEP peuvent également héberger et exprimer ce virus. Par contre, un témoin (donneur CTS) apparemment sain, possède des anticorps anti-gag (clone 2, 1077 pb). Ce qui est compatible avec une immunité acquise contre MSRV-1 en dehors d'une maladie autoimmune associée déclarée.

**20 EXEMPLE 7:** OBTENTION D'UN CLONE LB13 CONTENANT EN 3' UNE PARTIE HOMOLOGUE AU CLONE 2 CORRESPONDANT AU GENE GAG ET EN 5' UNE PARTIE HOMOLOGUE AU CLONE 5M6 CORRESPONDANT À LA RÉGION LTR U5.

25 Une RT-PCR ("reverse-transcriptase-polymerase chain reaction") a été effectuée à partir de l'ARN total extrait de virions, provenant de surnageants de cellules lymphocytaires B des patients atteints de sclérose en plaques, concentrés par ultracentrifugations. La synthèse de cDNA a été réalisée avec une amorce spécifique SEQ N° XXX et la transcriptase inverse Expand<sup>TM</sup> RT de BOEHRINGER MANNHEIM selon les conditions préconisées par la société.

Amorce utilisée pour la synthèse du cDNA, identifiée par

35 SEQ ID NO: 138:

5' CTT GGA GGG TGC ATA ACC AGG GAA T 3'

Une amplification par PCR a été réalisée avec la *Taq* polymérase (Perkin Elmer, France) sous les conditions suivantes: 94°C 1 min, 55°C 1 min, 72°C 2 min pendant 35 cycles et 72°C pendant 7 min et avec un volume réactionnel final de 100 µl.

Amorces utilisées pour l'amplification par PCR:

- amorce 5', identifiée par SEQ ID NO: 139

5' TGT CCG CTG TGC TCC TGA TC 3'

- amorce 3', identifiée par SEQ ID NO: 138

10 5' CTT GGA GGG TGC ATA ACC AGG GAA T 3'

Une deuxième amplification par PCR dite "semi-nichée" a été réalisée avec une amorce 3' située à l'intérieur de la région déjà amplifiée. Cette deuxième amplification a été effectuée sous les mêmes conditions expérimentales que celles utilisées lors de la première amplification, en utilisant 10 µl du produit d'amplification issu de la première PCR.

Amorces utilisées pour l'amplification par PCR "semi-nichée":

20 - amorce 5', identifiée par SEQ ID NO: 139

5' TGT CCG CTG TGC TCC TGA TC 3'

- amorce 3', identifiée par SEQ ID NO: 140

5' CTA TGT CCT TTT GGA CTG TTT GGG T 3'

Les amorces SEQ ID NO: 138 et SEQ ID NO: 140 sont spécifiques de la région gag, clone 2 position nucléotidique n° 373-397 et n° 433-456. Les amorces utilisées dans la région 5' ont été définies sur des séquences des clones obtenus lors d'essais préalables.

Un produit d'amplification de 764 pb a été obtenu et cloné de la façon suivante:

L'ADN amplifié a été inséré dans un plasmide à l'aide du kit TA Cloning<sup>TM</sup>. Les 2 µl de solution d'ADN ont été mélangés avec 5 µl d'eau distillée stérile, 1 µl d'un tampon de ligation 10 fois concentré "10X LIGATION BUFFER", 2 µl de "PCR<sup>TM</sup> VECTOR" (25 ng/ml) et 1 µl de "T4 DNA LIGASE". Ce mélange a été incubé une nuit à 14°C. Les

étapes suivantes ont été réalisées conformément aux instructions du kit TA Cloning® (Invitrogen). Le mélange a été étalé après transformation de la ligation dans des bactéries *E. coli* INVαF'. A la fin de la procédure, les colonies blanches de bactéries recombinantes ont été repiquées pour être cultivées et permettre l'extraction des plasmides incorporés selon la procédure dite de "minipréparation d'ADN" (17). La préparation de plasmide de chaque colonie recombinante a été coupée par l'enzyme de restriction *Eco* RI et analysée sur gel d'agarose. Les plasmides possédant un insert détecté sous lumière UV après marquage du gel au bromure d'éthidium, ont été sélectionnés pour le séquençage de l'insert, après hybridation avec une amorce complémentaire du promoteur T7 présent sur le plasmide de clonage du TA cloning kit®. La réaction préalable au séquençage a ensuite été effectuée selon la méthode préconisée pour l'utilisation du kit de séquençage "PRISM<sup>TM</sup> Ready Reaction Amplitaq® FS, DyeDeoxy<sup>TM</sup> Terminator" (Applied Biosystems, ref. 402119) et le séquençage automatique a été réalisé sur les appareils 373 A et 377 Applied Biosystems, selon les instructions du fabricant.

Le clone LB13 obtenu contient une région N-terminale de gène gag MSRV-1 homologue au clone 2 et un LTR correspondant à une partie de la région U5. Entre la région U5 et gag un site de fixation pour les ARN de transfert, le PBS "primer binding site" a été identifié.

La séquence nucléotidique du fragment de 764 pb du clone LB13 dans le plasmide "PCR<sup>TM</sup> vector" est représentée dans l'identificateur SEQ ID NO: 141.

Le site de fixation pour les ARN de transfert, présentant une séquence du type PBS tryptophane, a été identifié en position nucléotidique n°342-359 du clone LB13.

Comme ce même PBS a été retrouvé dans les copies endogènes homologues à MSRV1, la famille endogène ainsi

définie est appelée dorénavant HERV W, selon la nomenclature proposée pour les familles de rétrovirus endogènes (W=Tryptophane).

Une ORF courte d'environ 65 acides aminés a été retrouvée dans la région U5 du LTR 5' du clone LB13.

Séquence de l'ORF :

PMASNRAITLTAWSKIPFLGIRETKNPRSENTRLATMLEAAHHHFGSSPPLSWELWEQ  
GPQVTIW.

10 La séquence nucléotidique correspondante débutant par un codon ATG est susceptible d'être exprimée dans un ADN sous génomique à partir d'un LTR proviral (U3RU5).

Un autre clone, dénommé LA15 a été obtenu sur 15 l'ARN total extrait de virions concentrés par ultracentrifugation à partir d'un surnageant de culture de synoviocytes provenant d'un patient atteint de polyarthrite rhumatoïde. La stratégie d'amplification et clonage du clone LA15 est exactement la même qui a été 20 utilisée pour le clone LB13.

La séquence nucléotidique du clone LA15 qui est représentée dans l'identificateur SEQ ID NO: 142, est très similaire au clone LB13. Ceci suggère que le rétrovirus MSVR-1 détecté dans la sclérose en plaque présente des 25 séquences similaires à celles rencontrées dans la polyarthrite rhumatoïde.

EXEMPLE 8: RECONSTRUCTION D'UNE REGION RU5-GAG A PARTIR DES CLONES LB15, LB13, CL2 ET CL17

30 Les clones CL2 et LB13 ont déjà été décrits dans les exemples précédents. Le clone LB15 a été obtenu en utilisant la séquence R du LTR du clone cl6 pour définir une amorce en 5' et les amorces anti-sens utilisées sont 35 les mêmes que pour le clone LB13. Le clone CL17 a été

obtenu par nested RT-PCR en utilisant les amorces suivantes :

- 5'-TCATGCAACTGCACTCTTCTGGTCCG-3' (sens)  
 5'-TCTTGCACTAACCTCCACTGTCCGTTGG-3' (anti-sens)  
 5'-ATCCCCCAGTAACAATTTGGTGACCACG-3' (sens)  
 5'-TCGGGTCTAAGAGGGTACTTCCTTGGTAGG-3' (anti-sens)

- 10 Le clone LB15 a été obtenu à partir de virions obtenus par culture de cellules de SEP. Le clone LB17 a été obtenu à partir de culture de plasma de patient SEP.

Ces clones chevauchants ont permis de reconstruire une séquence RU5-gag avec une ORF potentielle dans le gène gag, telles que présentées à la figure 14.

#### **EXEMPLE 9: OBTENTION D'UN CLONE 87-23**

La région correspondant à l'intégrase a été amplifiée et clonée à partir de plasma de SEP en utilisant une semi-nested RT PCR avec les amorces suivantes situées dans les régions pol et env de MSRVI.

Dans la région pol :

5'-TTACGCAGGTCTCAGGGATGAGCTT-3' (sens-PCR primaire)  
 5'-CGGCAGTAGCAGTCTTAGTATCTGAAGCAGTTA-3' (sens-PCR secondaire)

Dans la région env,

5'-GGTACGGAGGGTTTCATGTAGTTTTGAG-3' (anti-sens PCR primaire et secondaire)

Le clone amplifié comporte 774 pb dans la région pol/RT, toute la région intégrase (1197 pb) et le début de la région env (480 pb). La séquence nucléotidique correspondant à la région intégrase et la traduction en

acides aminés de l'ORF potentiel sont présentés à la figure 15.

**EXEMPLE 10:** CONFIRMATION DE LA PRESENCE D'ARN CONTENANT  
5 DES SEQUENCES ENV APPARENTEES A ERV9 DANS LES PARTICULES  
RETROVIRALES ASSOCIEES AU GENOME MSRV1 :

Des séquences apparentées à ERV9 ont été trouvées  
en proportion minoritaire dans les préparations du virion  
10 provenant de SEP par rapport aux séquences MSRV1.  
L'existence de phénomènes de co-encapsidation de séquences  
endogènes phylogéniquement proches dans des particules  
rétrovirales produites par une souche répliquative a été  
décrite. De manière surprenante une région d'ARN  
15 comprenant une ORF commençant dans la partie 3' d'env et  
se continuant potentiellement dans le LTR3' a été  
retrouvée dans différents échantillons de SEP. Afin de  
préciser l'existence d'une ORF des essais de  
transcription-traduction ont été réalisés et ont permis de  
20 montrer la réalité d'une ORF env contenant toute la partie  
transmembranaire (TM) et se terminant au début du LTR  
putatif. Cependant une trame additionnelle (ORFX) fait  
suite et se poursuit dans le LTR3'. Les deux produits  
d'expression ont été visualisés et leurs ORFs respectives  
25 ont été sous clonées. La figure 16 représente les  
séquences nucléotidique et peptidique du clone B13 déjà  
décrit en précisant les ORFs dans la région env tronquée  
et dans le LTR putatif. La présence de tels ARN peut être  
à l'origine de recombinaisons avec la souche répliquative  
30 et par conséquent générer des souches de pathogénécité  
modifiée.

## BIBLIOGRAPHIE

- (1) Laemmli U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. 5 Nature. (1970). 227: 680-685.
- (2) Towbin H., Staehelin T. & Gordon J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some 10 applications. Proc. Natl. Acad. Sci. USA. (1979). 76: 4350-4354.

## REVENDEICATIONS

1. Matériel nucléique, à l'état isolé ou purifié, comprenant une séquence nucléotidique choisie dans le  
5 groupe qui consiste en (i) les séquences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii) les séquences complémentaires des séquences (i) ; et (iii) les séquences équivalentes aux  
10 séquences (i) ou (ii), en particulier les séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins 70 % d'homologie avec respectivement les séquences (i) ou (ii).
2. Matériel nucléique, à l'état isolé ou purifié,  
15 codant pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminées, au moins 50 %, et de préférence au moins 70 % d'homologie, avec une séquence peptidique choisie dans le groupe qui consiste en  
SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118,  
20 SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137.
3. Matériel nucléique rétroviral, dont le gène pol comprend une séquence nucléotidique identique ou équivalente à une séquence choisie dans le groupe qui  
consiste en SEQ ID NO: 112, SEQ ID NO: 124 et leurs  
25 séquences complémentaires.
4. Matériel nucléique rétroviral, dont l'extrémité  
5' du gène pol commence au nucléotide 1419 de  
SEQ ID NO: 130.
5. Matériel nucléique rétroviral, dont le gène pol  
30 code pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminées, au moins 50 %, et de préférence au moins 70 % d'homologie, avec la séquence peptidique SEQ ID NO: 113.
6. Matériel nucléique rétroviral, dont l'extrémité  
35 3' du gène gag finit au nucléotide 1418 de SEQ ID NO: 130.



7. Matériel nucléaire rétroviral, dont le gène env comprend une séquence nucléotidique identique ou équivalente à une séquence choisie dans le groupe qui consiste en SEQ ID NO: 117, et ses séquences  
5 complémentaires.

8. Matériel nucléaire rétroviral, dont le gène env comprend une séquence nucléotidique qui commence au nucléotide 1 de SEQ ID NO: 117 et finit au nucléotide au nucléotide 233 de SEQ ID NO: 114.

10 9. Matériel nucléaire rétroviral, dont le gène env code pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec la séquence SEQ ID NO: 118.

15 10. Matériel nucléaire rétroviral dont la région U3R du LTR 3' comprend une séquence nucléotidique qui se termine au nucléotide 617 de SEQ ID NO: 114.

11. Matériel nucléaire rétroviral dont la région RU5' du LTR 5' comprend une séquence nucléotidique qui  
20 commence au nucléotide 755 de SEQ ID NO: 120 et finit au nucléotide 337 de SEQ ID NO: 141 ou SEQ ID NO: 142.

12. Matériel nucléaire rétroviral comprenant une séquence qui commence au nucléotide 755 de SEQ ID NO: 120 et qui se termine au nucléotide 617 de SEQ ID NO: 114.

25 13. Matériel nucléaire rétroviral selon l'une quelconque des revendications précédentes, caractérisé en ce qu'il est associé à au moins une maladie auto-immune telle que la sclérose en plaques ou la polyarthrite rhumatoïde.

30 14. Fragment nucléotidique comprenant une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii)  
35 les séquences complémentaires des séquences (i) ; et (iii) les séquences équivalentes aux séquences (i) ou (ii), en

particulier les séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins 70 % d'homologie avec respectivement les séquences (i) ou (ii).

5 15. Fragment nucléotidique selon la revendication 14, consistant en une séquence nucléotidique choisie dans le groupe qui consiste en (i) les séquences SEQ ID NO: 112, SEQ ID NO: 114, SEQ ID NO: 117, SEQ ID NO: 120, SEQ ID NO: 124, SEQ ID NO: 130, 10 SEQ ID NO: 141 et SEQ ID NO: 142 ; (ii) les séquences complémentaires des séquences (i) ; et (iii) les séquences équivalentes aux séquences (i) ou (ii), en particulier les séquences présentant pour toute suite de 100 monomères contigus, au moins 50 %, et préférentiellement au moins 15 70 % d'homologie avec respectivement les séquences (i) ou (ii).

16. Fragment nucléotidique comprenant une séquence nucléotidique codant pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au 20 moins 50 %, et de préférence au moins 70 % d'homologie, avec une séquence peptidique choisie dans le groupe qui consiste en SEQ ID NO: 113, SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137.

25 17. Fragment nucléotidique selon la revendication 16, consistant en une séquence nucléotidique codant pour un polypeptide présentant, pour toute suite contigue d'au moins 30 acides aminés, au moins 50 %, et de préférence au moins 70 % d'homologie, avec une séquence peptidique choisie dans le groupe qui consiste en SEQ ID NO: 113, 30 SEQ ID NO: 115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137.

18. Sonde nucléique pour la détection d'un 35 rétrovirus associé à la sclérose en plaques et/ou la polyarthrite rhumatoïde, caractérisée en ce qu'elle est susceptible de s'hybrider spécifiquement sur tout fragment

sa selon l'une quelconque des revendications 14 à 17, appartenant au génome dudit rétrovirus.

19. Sonde selon la revendication 18, caractérisée en ce qu'elle possède de 10 à 100 nucléotides, de préférence de 10 à 30 nucléotides.

20. Amorce pour l'amplification par polymérisation d'un ARN ou d'un ADN d'un rétrovirus associé à la sclérose en plaques et/ou la polyarthrite rhumatoïde, caractérisée en ce qu'elle comprend une séquence nucléotidique identique ou équivalente à au moins une partie de la séquence nucléotidique d'un fragment selon l'une quelconque des revendications 8 à 11, notamment une séquence nucléotidique présentant pour toute suite de 10 monomères contigus, au moins 50 %, de préférence au moins 70 % d'homologie avec au moins ladite partie dudit fragment.

21. Amorce selon la revendication 20, caractérisée en ce que sa séquence nucléotidique est choisie parmi SEQ ID NO: 116, SEQ ID NO: 119, SEQ ID NO: 122, SEQ ID NO: 123, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 132, et SEQ ID NO: 133.

22. ARN ou ADN, et notamment vecteur de réplication et/ou d'expression, comprenant un fragment génomique du matériel nucléique selon l'une quelconque des revendications 1 à 7 ou un fragment selon l'une quelconque des revendications 14 à 17.

23. Peptide codé par tout cadre de lecture ouvert appartenant à un fragment nucléotidique selon l'une quelconque des revendications 14 à 17, notamment polypeptide, par exemple oligopeptide formant ou comprenant un déterminant antigénique reconnu par des sérums de patients infectés par le virus MSRV-1, et/ou chez lesquels le virus MSRV-1 a été réactivé.

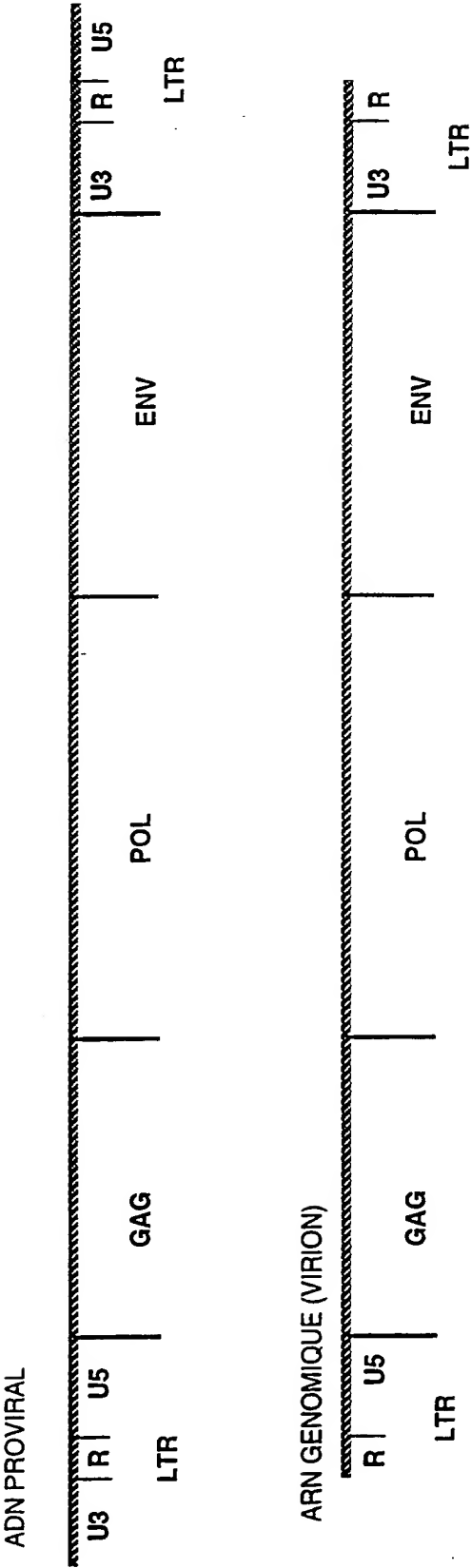
24. Peptide selon la revendication 23 comprenant une séquence identique, partiellement ou totalement, ou

équivalente à une séquence choisie parmi SEQ ID NO: 113, SEQ ID N°115, SEQ ID NO: 118, SEQ ID NO: 121, SEQ ID NO: 135 et SEQ ID NO: 137.

25. Composition diagnostique, prophylactique, ou  
5 thérapeutique, notamment pour inhiber l'expression d'au moins un rétrovirus associé à la sclérose en plaques et/ou à la polyarthrite rhumatoïde, comprenant un fragment nucléotidique selon l'une quelconque des revendications 14 à 17.

10 26. Procédé pour détecter un rétrovirus associé à la sclérose en plaques et/ou à la polyarthrite rhumatoïde, dans un échantillon biologique, caractérisé en ce que l'on met en contact un ARN et/ou un ADN présumé appartenir ou  
provenant dudit rétrovirus, ou leur ARN et/ou ADN  
15 complémentaire, avec une composition comprenant un fragment nucléotidique selon l'une quelconque des revendications 14 à 17.

1 / 32  
FIG 1



2 / 32

## FIG 2

| 10                                                      | 20            | 30          | 40         | 50         |     |
|---------------------------------------------------------|---------------|-------------|------------|------------|-----|
| 1234567890                                              | 1234567890    | 1234567890  | 1234567890 | 1234567890 |     |
| GCTTATAGAA GGACCCCTAG TATGGGGTAA TCCCTCTGG GAAACCAAGC   |               |             |            |            | 50  |
| A Y R R                                                 | T P S M G .   | S P L G     | N Q A      |            |     |
| L I E                                                   | G P L V       | W G N       | P L W      | E T K P    |     |
| L . K                                                   | D P .         | Y G V I     | P S G      | K P S      |     |
| CCCAGTACTC AGCAGGAAAA ATAGAATAGG AAACCTCACA AGGACATACT  |               |             |            |            | 100 |
| P V L                                                   | S R K N       | R I G       | N L T      | R T Y F    |     |
| Q Y S                                                   | A G K I E .   | E T S Q     | G H T      |            |     |
| P S T Q                                                 | Q E K .       | N R K P H K | D I L      |            |     |
| TTCTTCCCTT CCAGATGGCT AGCCACTGAG GAAGGAAAA TACTTTCACC   |               |             |            |            | 150 |
| P P L                                                   | Q M A S H .   | G R K N     | T F T      |            |     |
| F L P S                                                 | R W L A T E   | E G K I     | L S P      |            |     |
| S S P                                                   | P D G .       | P L R K E K | Y F H L    |            |     |
| TGCAGCTAAC CAACAGAAAT TACTTAAAAC CCTTCACCAA ACCTTCCACT  |               |             |            |            | 200 |
| C S .                                                   | P T E I T .   | N P S P N   | L P L      |            |     |
| A A N                                                   | Q Q K L       | L K T L H Q | T F H L    |            |     |
| Q L T                                                   | N R N Y L K P | F T K       | P S T      |            |     |
| TAGGCATIGA TAGCACCCAT CAGATGGCCA AATTATTATT TACTGGACCA  |               |             |            |            | 250 |
| R H .                                                   | . H P S       | D G Q I I I | Y W T R    |            |     |
| G I D                                                   | S T H Q M A K | L L F       | T G P      |            |     |
| . A L I                                                 | A P I R W P   | N Y Y L     | L D Q      |            |     |
| GGCTTTTICA AAACATATCAA GAAGATAGTC AGGGGCTGTG AAGTGTGCCA |               |             |            |            | 300 |
| P F Q                                                   | N Y Q E D S Q | G L .       | S V P      |            |     |
| G L F K                                                 | T I K K I V   | R G C E     | V C Q      |            |     |
| A F S                                                   | K L S R R .   | S G A V     | K C A K    |            |     |
| AAGAAATAAT                                              |               |             |            |            | 310 |
| K K .                                                   |               |             |            |            |     |
| R N N                                                   |               |             |            |            |     |
| E I                                                     |               |             |            |            |     |

3 / 32  
**FIG 2 (suite)**

| 10         | 20         | 30         | 40         | 50         |     |
|------------|------------|------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |     |
| CCCTGTATCT | TAAACCTCCT | TGTTAAGTTT | GTCTCTTCCA | GAATCAAAC  | 50  |
| P C I F    | N L L      | V K F      | V S S R    | I K T      |     |
| P V S      | L T S L    | L S L      | S L P      | E S K L    |     |
| L Y L      | . P P      | C . V C    | L F Q      | N Q N      |     |
|            |            |            |            |            |     |
| TGTAATACTA | CAAATTGTTC | TTCAAATGGA | GCACCAGATG | GAGTCCATGA | 100 |
| V K L      | Q I V L    | Q M E      | H Q M      | E S M T    |     |
| . N Y      | K L F      | F K W S    | T R W      | S P .      |     |
| C K T T    | N C S      | S N G      | A P D G    | V H D      |     |
|            |            |            |            |            |     |
| CTAAGATCCA | CCGTGGACCC | CTGGACCGGC | CTGCTAGCCC | ATGCTCCGAT | 150 |
| K I H      | R G P      | L D R P    | A S P      | C S D      |     |
| L R S T    | V D P      | W T G      | L L A H    | A P M      |     |
| . D P      | P W T P    | G P A      | C . P      | M L R C    |     |
|            |            |            |            |            |     |
| GTTAATGACA | TTGAAGGCAC | CCCTCCCGAG | GAAATCTCAA | CTGCACAACC | 200 |
| V N D I    | E G T      | P P E      | E I S T    | A Q P      |     |
| L M T      | L K A P    | L P R      | K S Q      | L H N P    |     |
| . . H      | . R H      | P S R G    | N L N      | C T T      |     |
|            |            |            |            |            |     |
| CCTACTATGC | CCCAATTCAG | CGGGAAGCAG | TTAGAGCGGT | CATCAGCCAA | 250 |
| L L C      | P N S A    | G S S      | . S G      | H Q P T    |     |
| Y Y A      | P I Q      | R E A V    | R A V      | I S Q      |     |
| P T M P    | Q F S      | G K Q      | L E R S    | S A N      |     |
|            |            |            |            |            |     |
| CCTCCCCAAC | AGCACTTGGG | TTTTCCTGTT | GAGAGGGGGG | ACTGAGAGAC | 300 |
| S P T      | A L G      | F S C      | . E G G    | L R D      |     |
| P P Q Q    | H L G      | F P V      | E R G D    | . E T      |     |
| L P N      | S T W V    | F L L      | R G G      | T E R Q    |     |
|            |            |            |            |            |     |
| AGGACTAGCT | GGATTTCCTA | GGCCAACGAA | GAATCCCTAA | GCCTAGCTGG | 350 |
| R T S W    | I S .      | A N E      | E S L S    | L A G      |     |
| G L A      | G F P R    | P T K      | N P .      | A . L G    |     |
| D . L      | D F L      | G Q R R    | I P K      | P S W      |     |

4 / 32  
FIG 3

| 10            | 20            | 30         | 40         | 50         |     |
|---------------|---------------|------------|------------|------------|-----|
| 1234567890    | 1234567890    | 1234567890 | 1234567890 | 1234567890 |     |
| GAAGGTGACT    | GCATCCACCT    | CTAAACATGG | GGCTTGCAAC | TTAGCTCACA | 400 |
| K V T A S T S | K H G A C N   | L A H T    |            |            |     |
| R . L H P P   | L N M G       | L A T .    | L T        |            |     |
| E G D C I H L | . T W G L Q L | S S H      |            |            |     |
| CCCGACCAAT    | CAGAGAGCTC    | ACTAAAATGC | TAATTAGGCA | AAAATAGGAG | 450 |
| R P I R E L   | T K M L I R Q | K . E      |            |            |     |
| P D Q S E S S | L K C . L G K | N R R      |            |            |     |
| P T N Q R A H | . N A N . A   | K I G G    |            |            |     |
| GTAAAGAAAT    | AGCCAATCAT    | CTATTGCCTG | AGAGCACAGC | GGGAGGGACA | 500 |
| V K K . P I I | Y C L R A Q R | E G Q      |            |            |     |
| . R N S Q S S | I A . E H S   | G R D K    |            |            |     |
| K E I A N H   | L L P E S T A | G G T      |            |            |     |
| AGGATCGGGA    | TATAAACCCA    | GGCATTGAG  | COGGCAACGG | CAACCCOCTT | 550 |
| G S G Y K P R | H S S R Q R   | Q P P L    |            |            |     |
| D R D I N P   | G I R A G N G | N P L      |            |            |     |
| R I G I . T Q | A F E P A T A | T P F      |            |            |     |
| TGGGTCCOCT    | CCCTTGTAT     | GGGCGCTCTG | TTTTCACTCT | ATTTCACTCT | 600 |
| G P L P L Y   | G R S V F T L | F H S      |            |            |     |
| W V P S L C M | G A L F S L Y | F T L      |            |            |     |
| G S P P F V W | A L C F H S   | I S L Y    |            |            |     |
| ATTAAATCTT    | GCAACTGAAA    | AAAAAAAAAA | AAAAA      |            | 635 |
| I K S C N . K | K K K K       |            |            |            |     |
| L N L A T E K | K K K K       |            |            |            |     |
| . I L Q L K   | K K K K       | K          |            |            |     |



5 / 32  
FIG 4

| 10         | 20         | 30          | 40         | 50         |     |
|------------|------------|-------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890  | 1234567890 | 1234567890 |     |
| ATGGCCCTCC | CTTATCATA  | CTTTCTCTTT  | ACTGTTCTCT | TACCCCTTT  | 50  |
| M A L P    | Y H T      | F L F       | T V L L    | P P F      |     |
| W P S      | L I I L    | F S L       | L F S      | Y P L S    |     |
| G P P      | L S Y      | F S L Y     | C S L      | T P F      |     |
| CGCTCTCACT | GCACCCCTC  | CATGCTGCTG  | TACAACCAGT | AGCTCCCTTT | 100 |
| A L T      | A P P P    | C C C       | T T S      | S S P Y    |     |
| L S L      | H P L      | H A A V     | Q P V      | A P L      |     |
| R S H C    | T P S      | M L L       | Y N Q      | . L P L    |     |
| ACCAAGAGTT | TCTATGAAGA | ACGCGGCTTC  | CTGGAATAT  | TGATGCCCCA | 150 |
| Q E F      | L . R      | T R L P     | G N I      | D A P      |     |
| T K S F    | Y E E      | R G F       | L E I L    | M P H      |     |
| P R V      | S M K N    | A A S       | W K Y      | . C P I    |     |
| TCATATAGGA | GTTTATCTAA | GGGAAACTCC  | ACCTTCACTG | CCCACACCCA | 200 |
| S Y R S    | L S K      | G N S       | T F T A    | H T H      |     |
| H I G      | V Y L R    | E T P       | P S L      | P T P I    |     |
| I . E      | F I .      | G K L H     | L H C      | P H P      |     |
| TATGCCCCGC | AACTGCTATA | ACTCTGCCAC  | TCTTTGCATG | CATGCAAATA | 250 |
| M P R      | N C Y N    | S A T       | L C M      | H A N T    |     |
| C P A      | T A I      | T L P L     | F A C      | M Q I      |     |
| Y A P Q    | L L .      | L C H       | S L H A    | C K Y      |     |
| CTCATTATIG | GACAGGGAAA | ATGATTAAATC | CTAGTTGTCC | TGGAGGACTT | 300 |
| H Y W      | T G K      | M I N P     | S C P      | G G L      |     |
| L I I G    | Q G K      | . L I       | L V V L    | E D L      |     |
| S L L      | D R E N    | D . S       | . L S      | W R T W    |     |
| GGAGCCACTG | TCTGTTGGAC | TCACTTCACC  | CATACCAGTA | TGCTTGATGG | 350 |
| G A T V    | C W T      | Y F T       | H T S M    | S D G      |     |
| E P L      | S V G L    | T S P       | I P V      | C L M G    |     |
| S H C      | L L D      | L L H P     | Y Q Y      | V . W      |     |

6 / 32  
**FIG 4 (suite)**

| 10         | 20          | 30         | 40         | 50          |      |
|------------|-------------|------------|------------|-------------|------|
| 1234567890 | 1234567890  | 1234567890 | 1234567890 | 1234567890  |      |
| ACCTCACCTG | TGTAAAATTT  | AGCAATACTA | TAGACACAAC | CAGCTCCCAA  | 750  |
| L T C      | V K F       | S N T I    | D T T      | S S Q       |      |
| T S P V    | . N L       | A I L      | . T Q P    | A P N       |      |
| P H L      | C K I       | . Q Y Y    | R H N      | Q L P M     |      |
| TGCATCAGGT | GGGTAAACACC | TCCCACACGA | ATAGTCTGCC | TACCCCTCAGG | 800  |
| C I R W    | V T P       | P T R      | I V C L    | P S G       |      |
| A S G      | G . H L     | P H E      | . S A      | Y P Q E     |      |
| H Q V      | G N T       | S H T N    | S L P      | T L R       |      |
| AATATTTTTT | GTCTGIGGTA  | CCTCAGCCTA | TCATTGTTTG | AATGGCTCTT  | 850  |
| I F F      | V C G T     | S A Y      | H C L      | N G S S     |      |
| Y F L      | S V V       | P Q P I    | I V .      | M A L       |      |
| N I F C    | L W Y       | L S L      | S L F E    | W L F       |      |
| CAGAATCTAT | GIGCTTCCTC  | TCATTCTTAG | TGCCCCCTAT | GACCATCTAC  | 900  |
| E S M      | C F L       | S F L V    | P P M      | T I Y       |      |
| Q N L C    | A S S       | H S .      | C P L .    | P S T       |      |
| R I Y      | V L P L     | I L S      | A P Y      | D H L H     |      |
| ACTGAACAAG | ATTATACAA   | TCATGTGGTA | CCTAAGCCCC | ACAACAAAAG  | 950  |
| T E Q D    | L Y N       | H V V      | P K P H    | N K R       |      |
| L N K      | I Y T I     | M S Y      | L S P      | T T K E     |      |
| . T R      | F I Q       | S C R T    | . A P      | Q Q K       |      |
| AGTACCCATT | CTTCCTTTTG  | TTATCAGAGC | AGGAGTGCTA | GGCAGACTAG  | 1000 |
| V P I      | L P F V     | I R A      | G V L      | G R L G     |      |
| Y P F      | F L L       | L S E Q    | E C .      | A D .       |      |
| S T H S    | S F C       | Y Q S      | R S A R    | Q T R       |      |
| GTACTGGCAT | TGGCAGTATC  | ACAACCTCTA | CTCAGTTCTA | CTACAAACTA  | 1050 |
| T G I      | G S I       | T T S T    | Q F Y      | Y K L       |      |
| V L A L    | A V S       | Q P L      | L S S T    | T N Y       |      |
| Y W H      | W Q Y H     | N L Y      | S V L      | L Q T I     |      |

FIG 4 (suite)

| 10         | 20          | 30         | 40          | 50          |      |
|------------|-------------|------------|-------------|-------------|------|
| 1234567890 | 1234567890  | 1234567890 | 1234567890  | 1234567890  |      |
| TCTCAAGAAA | TAAATGGTGA  | CATGGAACAG | GTCACITGACT | CCCTGGTTCAC | 1100 |
| S Q E I    | N G D       | M E Q      | V T D S     | L V T       |      |
| L K K      | . M V T     | W N R      | S L T       | P W S P     |      |
| S R N      | K W .       | H G T G    | H . L       | P G H       |      |
| CTTGCAAGAT | CAACTTAACT  | CCCTAGCAGC | AGTAGTCCTT  | CAAAATCGAA  | 1150 |
| L Q D      | Q L N S     | L A A      | V V L       | Q N R R     |      |
| C K I      | N L T       | P . Q Q    | . S F       | K I E       |      |
| L A R S    | T . L       | P S S      | S S P S     | K S K       |      |
| GAGCTTTAGA | CTTGCTAACC  | GCCAAAAGAG | GGGGAACCTG  | TTTATTTTIA  | 1200 |
| A L D      | L L T       | A K R G    | G T C       | L F L       |      |
| E L .      | T C .       | P P K E    | G E P V     | Y F .       |      |
| S F R      | L A N R     | Q K R      | G N L       | F I F R     |      |
| GGAGAAGAAC | GCTGTATATTA | TGTTAATCAA | TCCAGAATTG  | TCACITGAGAA | 1250 |
| G E E R    | C Y Y       | V N Q      | S R I V     | T E K       |      |
| E K N      | A V I M     | L I N      | P E L       | S L R K     |      |
| R R T      | L L L       | C . S I    | Q N C       | H . E       |      |
| AGTTAAAGAA | ATTGAGATC   | GAATACAATG | TAGAGCAGAG  | GAGCTTCAAA  | 1300 |
| V K E      | I R D R     | I Q C      | R A E       | E L Q N     |      |
| L K K      | F E I       | E Y N V    | E Q R       | S F K       |      |
| S . R N    | S R S       | N T M      | . S R G     | A S K       |      |
| ACACCGAACG | CTGGGGCCTC  | CTCAGCCAAT | GGATGCCCTG  | GGTTCTCCCC  | 1350 |
| T E R      | W G L       | L S Q W    | M P W       | V L P       |      |
| T P N A    | G A S       | S A N      | G C P G     | F S P       |      |
| H R T      | L G P P     | Q P M      | D A L       | G S P L     |      |
| TTCTTAGGAC | CTCTAGCAGC  | TCTAATATTG | TTACTCCTCT  | TTGGACCCCTG | 1400 |
| F L G P    | L A A       | L I L      | L L L F     | G P C       |      |
| S . D      | L . Q L     | . Y C      | Y S S       | L D P V     |      |
| L R T      | S S S       | S N I V    | T P L       | W T L       |      |

8 / 32  
**FIG 4 (suite)**

---

| 10                                     | 20         | 30         | 40         | 50         |      |
|----------------------------------------|------------|------------|------------|------------|------|
| 1234567890                             | 1234567890 | 1234567890 | 1234567890 | 1234567890 |      |
| TATCTTTAAC                             | CTCCTTGTTA | AGTTTGICTC | TTCCAGAATT | GAAGCTGTAA | 1450 |
| I F N                                  | L L V K    | F V S      | S R I      | E A V K    |      |
| S L T                                  | S L L      | S L S L    | P E L      | K L .      |      |
| Y L .                                  | P P C .    | V C L      | F Q N .    | S C K      |      |
| <br>AGCTACAGAT GGTCTTACAA ATGGAACCCC A |            |            |            |            | 1481 |
| L Q M                                  | V L Q      | M E P      |            |            |      |
| S Y R W                                | S Y K      | W N P      |            |            |      |
| A T D                                  | G L T N    | G T P      |            |            |      |

9 / 32

## FIG 5

| 10          | 20          | 30          | 40         | 50         |     |
|-------------|-------------|-------------|------------|------------|-----|
| 1234567890  | 1234567890  | 1234567890  | 1234567890 | 1234567890 |     |
| TCAAAATCGA  | AGAGCTTTAG  | ACTTGCTAAC  | CGCCAAAAGA | GGGGGAACCT | 50  |
| S K S K     | S F R       | L A N       | R Q K R    | G N L      |     |
| Q N R       | R A L D     | L L T       | A K R      | G G T C    |     |
| K I E       | E L .       | T C .       | P P K E    | G E P      |     |
|             |             |             |            |            |     |
| GTTTATTTTT  | AGGGGAAGAA  | TGCTGTTAGT  | ATGTTAATCA | ATCTGGAATC | 100 |
| F I F       | R G R M     | L L V C .   | S I W N H  |            |     |
| L F L       | G E E C C . | Y - V N Q   | S G I      |            |     |
| V Y F .     | G K N       | A V S       | M L I N    | L E S      |     |
|             |             |             |            |            |     |
| ATTACTGAGA  | AAGTTAAAGA  | AATTIGAGAT  | CGAATATAAT | GTAGAGCAGA | 150 |
| Y . E S .   | R N L R S   | N I M .     | S R        |            |     |
| I T E K     | V K E I .   | D R I .     | C R A E    |            |     |
| L L R       | K L K K     | F E I       | E Y N      | V E Q R    |     |
|             |             |             |            |            |     |
| GGACCTTCAA  | AACACTGCAC  | CCTGGGGCOCT | CCTCAGCCAA | TGGATGCCCT | 200 |
| G P S K     | H C T       | L G P       | P Q P M    | D A L      |     |
| D L Q       | N T A P     | W G L       | L S Q      | W M P W    |     |
| T F K       | T L H       | P G A S     | S A N      | G C P      |     |
|             |             |             |            |            |     |
| GGACTCTCCC  | CTTCTTAGGA  | CCTCTAGCAG  | CTATAATATT | TTTACTCCTC | 250 |
| D S P       | L L R T     | S S S       | Y N I      | F T P L    |     |
| T L P       | F L G       | P L A A     | I I F      | L L L      |     |
| G L S P     | S . D L .   | Q L .       | Y F        | Y S S      |     |
|             |             |             |            |            |     |
| TTTGGACCOCT | GTATCTTCAA  | CTTCCTTGTT  | AAGTTTGICT | CTTCCAGAAT | 300 |
| W T L       | Y L Q       | L P C .     | V C L      | F Q N      |     |
| F G P C     | I F N       | F L V       | K F V S    | S R I      |     |
| L D P       | V S S T     | S L L       | S L S      | L P E L    |     |
|             |             |             |            |            |     |
| TGAAGCTGTA  | AAGCTACAAA  | TAGTTCTTCA  | AATGGAACCC | CAGATGCAGT | 350 |
| . S C K     | A T N       | S S S       | N G T P    | D A V      |     |
| E A V       | K L Q I     | V L Q       | M E P      | Q M Q S    |     |
| K L .       | S Y K .     | F F K       | W N P      | R C S      |     |

10 / 32  
**FIG 5 (suite)**

| 10         | 20         | 30         | 40         | 50          |     |
|------------|------------|------------|------------|-------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890  |     |
| CCATGACTAA | AATCTACCGT | GGACCCCTGG | ACCGGCCTGC | TAGACTATGC  | 400 |
| H D .      | N L P W    | T P G      | P A C      | . T M L     |     |
| M T K      | I Y R      | G P L D    | R P A      | R L C       |     |
| P . L K    | S T V      | D P W      | T G L L    | D Y A       |     |
| TCTGATGTTA | ATGACATTGA | AGTCACCCCT | CCCGAGGAAA | TCTCAACTGC  | 450 |
| . C .      | . H .      | S H P S    | R G N      | L N C       |     |
| S D V N    | D I E      | V T P      | P E E I    | S T A       |     |
| L M L      | M T L K    | S P L      | P R K      | S Q L H     |     |
| ACAACCCCTA | CTACACTCCA | ATTCAGTAGG | AAGCAGTTAG | AGCAGTTGTC  | 500 |
| T T P T    | T L Q      | F S R      | K Q L E    | Q L S       |     |
| Q P L      | L H S N    | S V G      | S S .      | S S C Q     |     |
| N P Y      | Y T P      | I Q .      | E A V R    | A V V       |     |
| AGCCAACCTC | CCCAACAGTA | CTTGGGTTTT | CCTGTTGAGA | GGGTGGACTG  | 550 |
| A N L      | P N S T    | W V F      | L L R      | G W T E     |     |
| P T S      | P T V      | L G F S    | C . E      | G G L       |     |
| S Q P P    | Q Q Y      | L G F      | P V E R    | V D .       |     |
| AGAGACAGGA | CTAGCTGGAT | TTCCTAGGCT | GACTAAGAAT | CCCNAAAGCCT | 600 |
| R Q D      | . L D      | F L G .    | L R I      | P K P       |     |
| R D R T    | S W I      | S . A      | D . E      | S X S L     |     |
| E T G      | L A G F    | P R L      | T K N      | P X A X     |     |
| ANCTGGGAAG | GTGACCGCAT | CCATCTTTAA | ACATGGGGCT | TGCAACTTAG  | 650 |
| X W E G    | D R I      | H L .      | T W G L    | Q L S       |     |
| X G K      | V T A S    | I F K      | H G A      | C N L A     |     |
| L G R      | . P H      | P S L N    | M G L      | A T .       |     |
| CTCACACCCG | ACCAATCAGA | GAGCTCACTA | AAATGCTAAT | CAGGCAAAAA  | 700 |
| S H P      | T N Q R    | A H .      | N A N      | Q A K T     |     |
| H T R      | P I R      | E L T K    | M L I      | R Q K       |     |
| L T P D    | Q S E      | S S L      | K C .      | S G K N     |     |

## FIG 5 (suite)

| 10            | 20            | 30          | 40         | 50         |      |
|---------------|---------------|-------------|------------|------------|------|
| 1234567890    | 1234567890    | 1234567890  | 1234567890 | 1234567890 |      |
| CAGGAGGTAA    | AGCAATAGCC    | AATCATCTAT  | TGCTGAGAG  | CACAGCGGGA | 750  |
| G G K A I A   | N H L L       | P E S T A G |            |            |      |
| Q E V K Q .   | P I I Y       | C L R A     | Q R E      |            |      |
| R R . S N S Q | S S I A .     | E H S G K   |            |            |      |
| AGGACAAGGA    | TTGGGATATA    | AACTCAGGCA  | TTCAAGCCAG | CAACAGCAAC | 800  |
| R T R I G I . | T Q A F K P A | T A T       |            |            |      |
| G Q G L G Y K | L R H S S Q   | Q Q Q P     |            |            |      |
| D K D W D I   | N S G I       | Q A S N S N |            |            |      |
| CCCCTTTGGG    | TCCCCGCCA     | TTGTATGGGA  | GCTCTGTTTT | CACTCTATTT | 850  |
| P F G S P P I | V W E L C F   | H S I S     |            |            |      |
| P L G P L P   | L Y G S S V F | T L F       |            |            |      |
| P L W V P S H | C M G A L F S | L Y F       |            |            |      |
| CACTCTATTA    | AATCATGCAA    | CTGCACTCTT  | CTGGTCCGIG | TTTTTTATGG | 900  |
| L Y . I M Q   | L H S S G P C | F L W       |            |            |      |
| H S I K S C N | C T L L V R V | F Y G       |            |            |      |
| T L L N H A T | A L F W S V   | F F M A     |            |            |      |
| CTCAAGCTGA    | GCTTTTGTTC    | GCCATCCACC  | ACTGCTGTTC | GCCACCGTCA | 950  |
| L K L S F C S | P S T T A V C | H R H       |            |            |      |
| S S . A F V R | H P P L L F   | A T V T     |            |            |      |
| Q A E L L F   | A I H H C C L | P P S       |            |            |      |
| CAGACCCGCT    | GCTGACTTCC    | ATCCCTTTGG  | ATCCAGCAGA | GTGTCCACTG | 1000 |
| R P A A D F H | P F G S S R   | V S T V     |            |            |      |
| D P L L T S   | I P L D P A E | C P L       |            |            |      |
| Q T R C . L P | S L W I Q Q S | V H C       |            |            |      |
| TGCTCTGAT     | CCAGCGAGGT    | ACCCATTGCC  | ACTCCCGATC | AGGCTAAAGG | 1050 |
| L L I Q R G   | T H C H S R S | G . R       |            |            |      |
| C S . S S E V | P I A T P D Q | A K G       |            |            |      |
| A P D P A R Y | P L P L P I   | R L K A     |            |            |      |

12 / 32  
**FIG 5 (suite)**

| 10         | 20         | 30         | 40         | 50         |      |
|------------|------------|------------|------------|------------|------|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |      |
| CTTGCCATTG | TTCTGTCATG | GCTAAGTGCC | TGGGTTTGTC | CTAATAGAAC | 1100 |
| L A I V    | P A W      | L S A      | W V C P    | N R T      |      |
| L P L      | F L H G    | . V P      | G F V      | L I E L    |      |
| C H C      | S C M      | A K C L    | G L S      | . . N      |      |
|            |            |            |            |            |      |
| TGAACACTGG | TCACTGGGTT | CCATGGTTCT | CTTCCATGAC | CCACGGCTTC | 1150 |
| E H W      | S L G S    | M V L      | F H D      | P R L L    |      |
| N T G      | H W V      | P W F S    | S M T      | H G F      |      |
| . T L V    | T G F      | H G S      | L P .      | P T A S    |      |
|            |            |            |            |            |      |
| TAATAGAGCT | ATAACACTCA | CCGCATGGCC | CAAGATTCCA | TTCTTTGGTA | 1200 |
| I E L      | . H S      | P H G P    | R F H      | S L V      |      |
| . . S Y    | N T H      | R M A      | Q D S I    | P W Y      |      |
| N R A      | I T L T    | A W P      | K I P      | F L G I    |      |
|            |            |            |            |            |      |
| TCTGTGAGGC | CAAGAACCCC | AGGTCAGAGA | ANGTGAGGCT | TGCCACCATT | 1250 |
| S V R P    | R T P      | G Q R X    | . G L      | P P F      |      |
| L . G      | Q E P Q    | V R E      | X E A      | C H H L    |      |
| C E A      | K N P      | R S E X    | V R L      | A T I      |      |
|            |            |            |            |            |      |
| TGGGAAGTGG | CCCACTGCCA | TTTTGGTAGC | GGCCCACCAC | CATCTTGGGA | 1300 |
| G K W      | P T A I    | L V A      | A H H      | H L G S    |      |
| G S G      | P L P      | F W .      | R P T T    | I L G      |      |
| W E V A    | H C H      | F G S      | G P P P    | S W E      |      |
|            |            |            |            |            |      |
| GCTGTGGGAG | CAAGGATCCC | CCAGTAAACA |            |            | 1329 |
| C G S      | K D P      | P V T      |            |            |      |
| A V G A    | R I P      | Q .        |            |            |      |
| L W E      | Q G S P    | S N        |            |            |      |



13 / 32

FIG 6

| 10         | 20          | 30          | 40         | 50         |     |
|------------|-------------|-------------|------------|------------|-----|
| 1234567890 | 1234567890  | 1234567890  | 1234567890 | 1234567890 |     |
| CCTAGAACGT | ATTCTGGAGA  | ATTGGGACCA  | ATGTGACACT | CAGACGCTAA | 50  |
| P R T Y    | S G E       | L G P       | M . H S    | D A K      |     |
| L E R      | I L E N     | W D Q       | C D T      | Q T L R    |     |
| . N V      | F W R       | I G T N     | V T L      | R R .      |     |
|            |             |             |            |            |     |
| GAAAGAAACG | ATTTATATTTC | TTCTGCAGTA  | CCGCGTGGCC | ACAATATCCT | 100 |
| K E T      | I Y I L     | L Q Y       | R L A      | T I S S    |     |
| K K R      | F I F       | F C S T     | A W P      | Q Y P      |     |
| E R N D    | L Y S       | S A V       | P P G H    | N I L      |     |
|            |             |             |            |            |     |
| CTTCAAGCGA | GAGAAACCTG  | GCCTTCTGAG  | GGAAGTATAA | ATTATAACAT | 150 |
| S R E      | R N L       | A S . G     | K Y K      | L . H      |     |
| L Q G R    | E T W       | L P E       | G S I N    | Y N I      |     |
| F K G      | E K P G     | F L R       | E V .      | I I T S    |     |
|            |             |             |            |            |     |
| CATCTTACAG | CTAGACCTCT  | TCTGTAGAAA  | GGAGGGCAAA | TGGAGTGAAG | 200 |
| H L T A    | R P L       | L . K       | G G Q M    | E . S      |     |
| I L Q      | L D L F     | C R K       | E G K      | W S E V    |     |
| S Y S      | . T S       | S V E R     | R A N      | G V K      |     |
|            |             |             |            |            |     |
| TGCCATATGT | GCAAACCTTC  | TTTTCATTTAA | GAGACAATC  | ACAATTATGT | 250 |
| A I C      | A N F L     | F I K       | R Q L      | T I M .    |     |
| P Y V      | Q T F       | F S L R     | D N S      | Q L C      |     |
| C H M C    | K L S       | F H .       | E T T H    | N Y V      |     |
|            |             |             |            |            |     |
| AAAAAGTGIG | GTTTATGCCC  | TACAGGAAGC  | CCTCAGAGTC | CACTTCCTTA | 300 |
| K V W      | F M P       | Y R K P     | S E S      | T S L      |     |
| K K C G    | L C P       | T G S       | P Q S P    | P P Y      |     |
| K S V      | V Y A L     | Q E A       | L R V      | H L P T    |     |
|            |             |             |            |            |     |
| CCCCAGCGTC | CCCTCCCCGA  | CTCCTTCCTC  | AACTAATAAG | GACCCCCCTT | 350 |
| P Q R P    | L P D       | S F L N . . | G P P F    |            |     |
| P S V      | P S P T     | P S S       | T N K      | D P P L    |     |
| P A S      | P P R       | L L P Q     | L I R      | T P L      |     |
|            |             |             |            |            |     |
| TAACCCAAAC | GGTCCAAAAG  | GAGATAGACA  | AAGGGGTAAA | CAATGAACCA | 400 |
| N P N      | G P K G     | D R Q       | R G K      | Q . T K    |     |
| T Q T      | V Q K       | E I D K     | G V N      | N E P      |     |
| . P K R    | S K R       | R . T       | K G .      | T M N Q    |     |
|            |             |             |            |            |     |
| AAGAGTGGCA | ATATTCCCCG  | ATTATGCCCC  | CTCCAAGCAG | TGAGAGGAGG | 450 |
| E C Q      | Y S P       | I M P P     | P S S      | E R R      |     |
| K S A N    | I P R       | L C P       | L Q A V    | R G G      |     |
| R V P      | I F P D     | Y A P       | S K Q      | . E E E    |     |
|            |             |             |            |            |     |
| AGAATTGGGC | CCAGCCAGAG  | TGCCGTGTACC | TTTTTCTCTC | TCAGACTTAA | 500 |
| R I R P    | S Q S       | A C T       | F F S L    | R L K      |     |
| E F G      | P A R V     | P V P       | F S L      | S D L K    |     |
| N S A      | Q P E       | C L Y L     | F L S      | Q T .      |     |

14 / 32  
**FIG 6 (suite)**

| 10         | 20          | 30         | 40         | 50            |      |
|------------|-------------|------------|------------|---------------|------|
| 1234567890 | 1234567890  | 1234567890 | 1234567890 | 1234567890    |      |
| AGCAAATTAA | AATAGACCTA  | GGTAAATTCT | CAGATAACCC | TGACGGCTAT    | 550  |
| A N .      | N R P R .   | I L R .    | P .        | R L Y         |      |
| Q I K      | I D L       | G K F S    | D N P      | D G Y         |      |
| S K L K .  | T .         | V N S      | Q I T L    | T A I         |      |
| ATTGATGTTT | TACAAGGGTT  | AGGACAATCC | TTTGATCTGA | CATGGAGAGA    | 600  |
| . C F      | T R V       | R T I L .  | S D        | M E R         |      |
| I D V L    | Q G L       | G Q S      | F D L T    | W R D         |      |
| L M F      | Y K G .     | D N P      | L I .      | H G E I       |      |
| TATAATGTTA | CTACTAAATC  | AGACACTAAC | CCCAAATGAG | AGAAGTGCCG    | 650  |
| Y N V T    | T K S       | D T N      | P K .      | E K C R       |      |
| I M L      | L L N Q     | T L T      | P N E      | R S A A       |      |
| . C Y Y .  | I R H .     | P          | Q M R      | E V P         |      |
| CTGTAAGTGC | AGCCCGAGAG  | TTTGGCGATC | TTTGGTATCT | CAGTCAGGCC    | 700  |
| C N C      | S P R V     | W R S      | L V S      | Q S G Q       |      |
| V T A      | A R E       | F G D L    | W Y L      | S Q A         |      |
| L .        | L Q         | P E S      | L A I      | F G I S V R P |      |
| AACAATAGGA | TGACAACAGA  | GGAAAGAACA | ACTCCACAG  | GCCAGCAGGC    | 750  |
| Q .        | D D N R     | G K N N    | S H R      | P A G         |      |
| N N R M    | T T E       | E R T      | T P T G    | Q Q A         |      |
| T I G .    | Q Q R       | K E Q      | L P Q      | A S R Q       |      |
| AGTTCACAGT | GTAGACCCCTC | ATTGGGACAC | AGAATCAGAA | CATGGAGATT    | 800  |
| S S Q C    | R P S       | L G H      | R I R T    | W R L         |      |
| V P S      | V D P H     | W D T      | E S E      | H G D W       |      |
| F P V .    | T L         | I G T Q    | N Q N      | M E I         |      |
| GGTGCCACAA | ACATTTCGTA  | ACTTGGGTGC | TAGAAGGACT | GAGGAAACT     | 850  |
| V P Q      | T F A N     | L R A      | R R T      | E E N .       |      |
| C H K      | H L L       | T C V L    | E G L      | R K T         |      |
| G A T N    | I C .       | L A C .    | K D .      | G K L         |      |
| AGGAAGAAGC | CTATGAATTA  | CTCAATGATG | TCCACTATAA | CACAGGGAAA    | 900  |
| E E A      | Y E L       | L N D V    | H Y N      | T G K         |      |
| R K K P    | M N Y       | S M M      | S T I T    | Q G K         |      |
| G R S      | L .         | I T Q .    | C P L .    | H R E R       |      |
| GGAAGAAAT  | CTTACTGCCT  | TCTGGACAG  | ACTAAGGGAG | GCATTGAGGA    | 950  |
| G R K S    | Y C F       | S G Q      | T K G G    | I E E         |      |
| E E N      | L T A F     | L D R      | L R E      | A L R K       |      |
| K K I      | L L L       | F W T D .  | G R H .    | G             |      |
| AGCATACCTC | CCTGTACCT   | GACTCTATTG | AAGGCCAACT | AATCTTAAAG    | 1000 |
| A Y L      | P V T .     | L Y .      | R P T      | N L K G       |      |
| H T S      | L S P       | D S I E    | G Q L      | I L K         |      |
| S I P P    | C H L       | T L L      | K A N .    | S . R         |      |

15 / 32  
FIG 6 (suite)

| 10                  | 20                  | 30          | 40         | 50         |      |
|---------------------|---------------------|-------------|------------|------------|------|
| 1234567890          | 1234567890          | 1234567890  | 1234567890 | 1234567890 |      |
| GATAAGTTTA          | TCACCTCAGTC         | AGCTGCAGAC  | ATTAGAAAAA | ACTTCAAAAG | 1050 |
| . V Y H S V         | S C R H             | . K K L Q K |            |            |      |
| D K F I T Q S       | A A D I R K N       | F K S       |            |            |      |
| I S L S L S Q       | L Q T L E K         | T S K V     |            |            |      |
| TCTGCGTTAG          | GCCCCGAGCA          | GAACCTTAGAA | ACCCTATTTA | ACTTGGCATC | 1100 |
| S A L G P E Q       | N L E T L F N       | L A S       |            |            |      |
| L P . A R S R T     | . K P Y L T W H P   |             |            |            |      |
| C L R P G A E L R N | P I . L G I         |             |            |            |      |
| CTCAGITTTT          | TATAATAGAG          | ATCAGGAGGA  | GCAGGCGAAA | CGGGACAAAC | 1150 |
| S V F Y N R D       | Q E E Q A K R D K R |             |            |            |      |
| Q F F I I E I R R S | R R N G T N         |             |            |            |      |
| L S F L . . R S G G | A G E T G Q T       |             |            |            |      |
| GGGATAAAAA          | AAAAAGGGGG          | GGTCCACTAC  | TTTAGTCATG | GCCCTCAGGC | 1200 |
| D K K K R G G P L L | . S W P S G         |             |            |            |      |
| G I K K K G G V H Y | F S H G P Q A       |             |            |            |      |
| G . K K K G G S T T | L V M A L R Q       |             |            |            |      |
| AAGCAGACTT          | TGGAGGCTCT          | GCAAAAGGGA  | AAAGCTGGGC | AAATCAAATG | 1250 |
| K Q T L E A L Q K G | K A G Q I K C       |             |            |            |      |
| S R L W R L C K R E | K L G K S N A       |             |            |            |      |
| A D F G G S A K G K | S W A N Q M         |             |            |            |      |
| OCTAATAGGG          | CTGGCTTCCA          | GTGGGGTCTA  | CAAGGACACT | TTAAAAAGA  | 1300 |
| L I G L A S S A V Y | K D T L K K I       |             |            |            |      |
| . . G W L P V R S T | R T L . K R         |             |            |            |      |
| P N R A G F Q C G L | Q G H F K K D       |             |            |            |      |
| TTATCCAAGT          | AGAAATAAGC          | CGCCCCCTTG  | TCCATGCCCC | TTAAGTCAAG | 1350 |
| I Q V E I S R P L V | H A P Y V K         |             |            |            |      |
| L S K . K . A A P L | S M P L T S R       |             |            |            |      |
| Y P S R N K P P P C | P C P L R Q G       |             |            |            |      |
| GGAATCACTG          | GAAGGCCCCAC         | TGCCCCAGGG  | GATCAAGATA | CTCTGAGTCA | 1400 |
| G I T G R P T A P G | D E D T L S Q       |             |            |            |      |
| E S L E G P L P Q G | M K I L . V R       |             |            |            |      |
| N H W K A H C P R G | . R Y S E S         |             |            |            |      |
| GAAGCCATTA          | ACCAGATGAT          | CCAGCAGCAG  | GACTGAGGGT | GGCCGGGGGG | 1450 |
| K P L T R . S S S R | T E G A R G E       |             |            |            |      |
| S H . P D D P A A G | L R V P G A         |             |            |            |      |
| E A I N Q M I Q Q Q | D . G C P G R       |             |            |            |      |
| AGCGGCAGCC          | CATGOCATCA          | COCTCACAGA  | GGCCGGGGTA | TGTTTGACCA | 1500 |
| R Q P M P S P S Q S | P G Y V . P         |             |            |            |      |
| S A S P C H H P H R | A P G M F D H       |             |            |            |      |
| A P A H A I T L T E | P R V C L T I       |             |            |            |      |

16 / 32  
**FIG 6 (suite)**

| 10           | 20         | 30         | 40         | 50         |
|--------------|------------|------------|------------|------------|
| 1234567890   | 1234567890 | 1234567890 | 1234567890 | 1234567890 |
| TTGAGAGCCA A |            |            |            | 1511       |
| L R A        |            |            |            |            |
| . E P        |            |            |            |            |
| E S Q        |            |            |            |            |

17 / 32

FIG 7

| 10         | 20         | 30         | 40          | 50         |     |
|------------|------------|------------|-------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890  | 1234567890 |     |
| ATGGGCAGCA | GCCATCATCA | TCATCATCAC | AGCAGCGGCC  | TGGTGCCGCG | 50  |
| M G S S    | H H H      | H H H      | S S G L     | V P R      |     |
| CGGCAGCCAT | ATGGCTAGCA | TGACTGGTGG | ACAGCAAATG  | GGTCGGATCC | 100 |
| G S H      | M A S M    | T G G      | Q Q M       | G R I L    |     |
| TAGAACTAT  | TCTGGAGAAT | TGGGACCAAT | GTGACACTCA  | GACGCTAAGA | 150 |
| E R I      | L E N      | W D Q C    | D T Q       | T L R      |     |
| AAGAAACGAT | TTATATTCTT | CTGCAGTACC | GCCTGGCCAC  | AATATCCTCT | 200 |
| K K R F    | I F F      | C S T      | A W P Q     | Y P L      |     |
| TCAAGGGAGA | GAAACCTGGC | TTCCTGAGGG | AAGTATAAAT  | TATAACATCA | 250 |
| Q G R      | E T W L    | P E G      | S I N       | Y N I I    |     |
| TCTTACAGCT | AGACCTCTTC | TGTAGAAAGG | AGGGCAAATG  | GAGTGAAGTG | 300 |
| L Q L      | D L F      | C R K E    | G K W       | S E V      |     |
| CCATATGTGC | AAACTTTCTT | TTCATTAAGA | GACAACTCAC  | AATTATGTAA | 350 |
| P Y V Q    | T F F      | S L R      | D N S Q     | L C K      |     |
| AAAGTGTGGT | TTATGCCCTA | CAGGAAGCCC | TCAGAGTCCA  | CCTCCCTACC | 400 |
| K C G      | L C P T    | G S P      | Q S P       | P P Y P    |     |
| CCAGGTCCC  | CTCCCCGACT | CCTTCCTCAA | CTAATAAGGA  | CCCCCCTTTA | 450 |
| S V P      | S P T      | P S S T    | N K D       | P P L      |     |
| ACCCAAACGG | TCCAAAAGGA | GATAGACAAA | GGGTAAACA   | ATGAACCAAA | 500 |
| T Q T V    | Q K E      | I D K      | G V N N     | E P K      |     |
| GAGTGCCAAT | ATTCCCCGAT | TATGCCCCCT | CCAAGCAGTG  | AGAGGAGGAG | 550 |
| S A N      | I P R L    | C P L      | Q A V       | R G G E    |     |
| AATTGGGCCC | AGCCAGAGTG | CCTGTACCTT | TTTCTCTCTC  | AGACTTAAAG | 600 |
| F G P      | A R V      | P V P F    | S L S       | D L K      |     |
| CAAATTAAAA | TAGACCTAGG | TAAATTCTCA | GATAACCCCTG | ACGGCTATAT | 650 |
| Q I K I    | D L G      | K F S      | D N P D     | G Y I      |     |
| TGATGTTTTA | CAAGGGTTAG | GACAATCCTT | TGATCTGACA  | TGGAGAGATA | 700 |
| D V L      | Q G L G    | Q S F      | D L T       | W R D I    |     |
| TAATGTTACT | ACTAAATCAG | ACACTAATCC | CAAATGAGAG  | AAGTGCCGCT | 750 |
| M L L      | L N Q      | T L T P    | N E R       | S A A      |     |
| GTAAGTGCAG | CCCGAGAGTT | TGGCGATCTT | TGGTATCTCA  | GTCAGGOCOA | 800 |
| V T A A    | R E F      | G D L      | W Y L S     | Q A N      |     |

18 / 32  
**FIG 7 (suite)**

| 10         | 20         | 30         | 40          | 50         |      |
|------------|------------|------------|-------------|------------|------|
| 1234567890 | 1234567890 | 1234567890 | 1234567890  | 1234567890 |      |
| CAATAGGATG | ACAACAGAGG | AAAGAACAAC | TCCCACAGGC  | CAGCAGGCAG | 850  |
| N R M      | T T E E    | R T T      | P T G       | Q Q A V    |      |
| TTCCCAGTGT | AGACCCTCAT | TGGGACACAG | AATCAGAACA  | TGGAGATTGG | 900  |
| P S V      | D P H      | W D T E    | S E H       | G D W      |      |
| TGCCACAAAC | ATTIGCTAAC | TTGGGTGCTA | GAAGGACTGA  | GGAAACTAG  | 950  |
| C H K H    | L L T      | C V L      | E G L R     | K T R      |      |
| GAAGAAGCCT | ATGAATTACT | CAATGATGTC | CACTATAACA  | CAGGGAAAGG | 1000 |
| K K P      | M N Y S    | M M S      | T I T       | Q G K E    |      |
| AAGAAAATCT | TACTGCTTTT | CTGGACAGAC | TAAGGGAGGC  | ATTGAGGAAG | 1050 |
| E N L      | T A F      | L D R L    | R E A       | L R K      |      |
| CATACCTCCC | TGTCACCTGA | CTCTATTGAA | GGCCAACTAA  | TCTTAAAGGA | 1100 |
| H T S L    | S P D      | S I E      | G Q L I     | L K D      |      |
| TAAGTTTATC | ACTCAGTCAG | CTGCAGACAT | TAGAAAAAAC  | TTCAAAAGTC | 1150 |
| K F I      | T Q S A    | A D I      | R K N       | F K S L    |      |
| TGCTTAAGCT | TGOGGCGCA  | CTCGAGCACC | ACCACCACCA  | CCACTGAGAT | 1200 |
| P K L      | A A A      | L E H H    | H H H       | H . D      |      |
| COGGCTGCTA | ACAAAGCCCG | AAAGGAAGCT | GAGTTGGCTIN | GTGGCNA    | 1247 |
| P A A N    | K A R      | K E A      | E L A X     | G          |      |

19 / 32  
FIG 8

| 10         | 20         | 30         | 40         | 50         |     |
|------------|------------|------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |     |
| ATGGCTAGCA | TGACTGGTGG | ACAGCAAATG | GGTCGGATCC | TAGAACGTAT | 50  |
| M A S M    | T G G      | Q Q M      | G R I L    | E R I      |     |
| TCTGGAGAAT | TGGGACCAAT | GTGACACTCA | GACGCTAAGA | AAGAAACGAT | 100 |
| L E N      | W D Q C    | D T Q      | T L R      | K K R F    |     |
| TTATATTCTT | CTGCAGTACC | GCCTGGCCAC | AATATCCTCT | TCAAGGGAGA | 150 |
| I F F      | C S T      | A W P Q    | Y P L      | Q G R      |     |
| GAAACCTGGC | TTCTTGAGGG | AAGTATAAAT | TATAACATCA | TCTTACAGCT | 200 |
| E T W L    | P E G      | S I N      | Y N I I    | L Q L      |     |
| AGACCTCTTC | TGTAGAAAGG | AGGGCAAATG | GAGTGAAGTG | CCATATGTGC | 250 |
| D L F      | C R K E    | G K W      | S E V      | P Y V Q    |     |
| AACTTTTCTT | TTCATTAAGA | GACAACTCAC | AATTATGTAA | AAAGTGTGGT | 300 |
| T F F      | S L R      | D N S Q    | L C K      | K C G      |     |
| TTATGCCCTA | CAGGAAGCCC | TCAGAGTCCA | CCTCCCTACC | CCAGCGTCCC | 350 |
| L C P T    | G S P      | Q S P      | P P Y P    | S V P      |     |
| CTCCCCGACT | CCTTCCTCAA | CTAATAAGGA | CCCCCTTTTA | ACCCAAACGG | 400 |
| S P T      | P S S T    | N K D      | P P L      | T Q T V    |     |
| TCAAAAGGA  | GATAGACAAA | GGGTAAACA  | ATGAACCAA  | GAGTGCCAAT | 450 |
| Q K E      | I D K      | G V N N    | E P K      | S A N      |     |
| ATTCCCCGAT | TATGCCCCCT | CCAAGCAGTG | AGAGGAGGAG | AATTCGGCCC | 500 |
| I P R L    | C P L      | Q A V      | R G G E    | F G P      |     |
| AGCAGAGTG  | CCTGTACCTT | TTTCTCTCTC | AGACTTAAAG | CAAATTAAAA | 550 |
| A R V      | P V P F    | S L S      | D L K      | Q I K I    |     |
| TAGACCTAGG | TAAATTCTCA | GATAACCCTG | ACGGCTATAT | TGATGTTTTA | 600 |
| D L G      | K F S      | D N P D    | G Y I      | D V L      |     |
| CAAGGGTTAG | GACAATCCTT | TGATCTGACA | TGGAGAGATA | TAATGTTACT | 650 |
| Q G L G    | Q S F      | D L T      | W R D I    | M L L      |     |
| ACTAAATCAG | ACACTAACC  | CAATGAGAG  | AAGTCCCGCT | GTAACGTCAG | 700 |
| L N Q      | T L T P    | N E R      | S A A      | V T A A    |     |
| CCCGAGAGTT | TGGCGATCTT | TGGTATCTCA | GTCAGGCCAA | CAATAGGATG | 750 |
| R E F      | G D L      | W Y L S    | Q A N      | N R M      |     |
| ACAACAGAGG | AAAGAACAAC | TCCACAGGC  | CAGCAGGCAG | TTCOCAGTGT | 800 |
| T T E E    | R T T      | P T G      | Q Q A V    | P S V      |     |

20 / 32

## FIG 8 (suite)

| 10          | 20         | 30         | 40         | 50         |      |
|-------------|------------|------------|------------|------------|------|
| 1234567890  | 1234567890 | 1234567890 | 1234567890 | 1234567890 |      |
| AGACCCATCAT | TGGGACACAG | AATCAGAACA | TGGAGATTGG | TGCCACAAAC | 850  |
| D P H       | W D T E    | S E H      | G D W      | C H K H    |      |
| ATTTGCTAAC  | TTGGGTGCTA | GAAGGACTGA | GGAAACTAG  | GAAGAAGCCT | 900  |
| L L T       | C V L      | E G L R    | K T R      | K K P      |      |
| ATGAATTACT  | CAATGATGTC | CACTATAACA | CAGGGAAAGG | AAGAAAATCT | 950  |
| M N Y S     | M M S      | T I T      | Q G K E    | E N L      |      |
| TACTGCTTTT  | CTGGACAGAC | TAAGGGAGGC | ATTGAGGAAG | CATACCTCCC | 1000 |
| T A F       | L D R L    | R E A      | L R K      | H T S L    |      |
| TGTCACCTGA  | CTCTATTGAA | GCCCAACTAA | TCTTAAAGGA | TAAGTTTATC | 1050 |
| S P D       | S I E      | G Q L I    | L K D      | K F I      |      |
| ACTCAGTCAG  | CTGCAGACAT | TAGAAAAAAC | TTCAAAAGTC | TGCTAAGCT  | 1100 |
| T Q S A     | A D I      | R K N      | F K S L    | P K L      |      |
| TGCGGCGCA   | CTCGAGCACC | ACCACCACCA | CCACTGAGAT | COGGCTGCTA | 1150 |
| A A A       | L E H H    | H H H      | H . D      | P A A N    |      |
| ACAAAGCCCCG | AAAGGAAGCT | GAGTTGGCTG | GTTGCA     |            | 1186 |
| K A R       | K E A      | E L A G    | G          |            |      |



FIG 9

| 10              | 20            | 30         | 40          | 50          |     |
|-----------------|---------------|------------|-------------|-------------|-----|
| 1234567890      | 1234567890    | 1234567890 | 1234567890  | 1234567890  |     |
| TGTCCTGCTGT     | GCTCCTGATC    | CAGCACAGGC | GCCCATTTGCC | TCTCCCAATT  | 50  |
| C P L C S . S   | S T G A H C L | S Q L      |             |             |     |
| V R C A P D P   | A Q A P I A   | S P N W    |             |             |     |
| S A V L L I     | Q H R R       | P L P      | L P I       |             |     |
| GGGCTAAAGG      | CTTGCCATTG    | TTCCTGCACA | GCTAAGTGGC  | TGGGTTTCATC | 100 |
| G . R L A I V   | P A Q L S A   | W V H P    |             |             |     |
| A K G L P L     | F L H S . V P | G F I      |             |             |     |
| G L K A C H C   | S C T A K C L | G S S      |             |             |     |
| CTAATCGAGC      | TGAACACTAG    | TCACTGGGTT | CCACGGTTCT  | CTTCCATGAC  | 150 |
| N R A E H .     | S L G S T V L | F H D      |             |             |     |
| L I E L N T S   | H W V P R F S | S M T      |             |             |     |
| . S S . T L V   | T G F H G S   | L P . P    |             |             |     |
| CCATGGCTTC      | TAATAGAGCT    | ATAACACTCA | CTGCATGGTC  | CAAGATTCCA  | 200 |
| P W L L I E L . | H S L H G P   | R F H      |             |             |     |
| H G F . . S Y   | N T H C M V   | Q D S I    |             |             |     |
| M A S N R A     | I T L T A W S | K I P      |             |             |     |
| TTCCTTGGA       | TCCGTGAGAC    | CAAGAACCCC | AGGTCAGAGA  | ACACAAGGCT  | 250 |
| S L E S V R P   | R T P G Q R   | T Q G L    |             |             |     |
| P W N P . D     | Q E P Q V R E | H K A      |             |             |     |
| F L G I R E T   | K N P R S E N | T R L      |             |             |     |
| TGCCACCATG      | TTGGAAGCAG    | CCCACCACCA | TTTTGGAAGC  | AGCCCCGCCAC | 300 |
| P P C W K Q     | P T T I L E A | A R H      |             |             |     |
| C H H V G S S   | P P P F W K Q | P A T      |             |             |     |
| A T M L E A A   | H H H F G S   | S P P L    |             |             |     |
| TATCTTGGGA      | GCTCTGGGAG    | CAAGGACCCC | AGGTAAACAAT | TTGGTGACCA  | 350 |
| Y L G S S G S   | K D P R . Q F | G D H      |             |             |     |
| I L G A L G A   | R T P G N N   | L V T T    |             |             |     |
| S W E L W E     | Q G P Q V T I | W . P      |             |             |     |
| CGAAGGGACC      | TGAATCCGCA    | ACCATGAAGG | GATCTCCAAA  | GCAATTGGAA  | 400 |
| E G T . I R N   | H E G I S K   | A I G N    |             |             |     |
| K G P . E S A   | T M K G S P K | Q L E      |             |             |     |
| R R D L N P Q   | P . R D L     | Q S N W K  |             |             |     |

22 / 32  
**FIG 9 (suite)**

| 10          | 20         | 30          | 40         | 50            |     |
|-------------|------------|-------------|------------|---------------|-----|
| 1234567890  | 1234567890 | 1234567890  | 1234567890 | 1234567890    |     |
| ATGTTCTCTCC | CAAGGCAAAA | ATGCCCCCTAA | GATGTATTCT | GGAGAATTGG    | 450 |
| V P P       | K A K      | M P L R     | C I L      | E N W         |     |
| M F L P     | R Q K      | C P .       | D V F W    | R I G         |     |
| C S S       | Q G K N    | A P K       | M Y S      | G E L G       |     |
| GACCAATTTG  | ACCCTCAGAC | AGTAAGAAAA  | AAATGACTTA | TATTCTTCTG    | 500 |
| D Q F D     | P Q T      | V R K K     | . L I      | F F C         |     |
| T N L       | T L R Q    | . E K       | N D L      | Y S S A       |     |
| P I .       | P S D      | S K K K     | M T Y      | I L L         |     |
| CAGTACCGCC  | CTGGCCACGA | TATCCTCTTC  | AAGGGGGAGA | AACCTGGCCT    | 550 |
| S T A       | L A T I    | S S S       | R G R      | N L A S       |     |
| V P P       | W P R      | Y P L Q     | G G E      | T W P         |     |
| Q Y R P     | G H D      | I L F       | K G E K    | P G L         |     |
| OCTGAGGGAA  | GTATAAATTA | TAACACCATC  | TTACAGCTAG | ACCTGTTTTG    | 600 |
| . G K       | Y K L      | . H H L     | T A R      | P V L         |     |
| P E G S     | I N Y      | N T I       | L Q L D    | L F C         |     |
| L R E V     | . I I      | T P S       | Y S .      | T C F V       |     |
| TAGAAAAGGA  | GGCAAATGGA | GIGAAGTGCC  | ATATTACAA  | ACITTCITTT    | 650 |
| . K R R     | Q M E      | . S A       | I F T N    | F L F         |     |
| R K G       | G K W S    | E V P       | Y L Q      | T F F S       |     |
| E K E       | A N G      | V K C H     | I Y K      | L S F         |     |
| CATTAAAAGA  | CAACTCGCAA | TTATGTTAAC  | AGTGTGATTT | GIGTTCCTAC    | 700 |
| I K R       | Q L A I    | M L T V     | . F        | V F L H       |     |
| L K D       | N S Q      | L C .       | Q          | C D L C S Y   |     |
| H .         | K T        | T R N       | Y V N      | S V I C V P T |     |
| ACGGAAGCCC  | TCAGATTCTA | CTCCCCACCC  | CCGGCATCTC | CCCTGAATCC    | 750 |
| G S P       | Q I L      | L P T P     | G I S      | P E S         |     |
| T E A L     | R F Y      | S P P       | P A S P    | L N P         |     |
| R K P       | S D S T    | P H P       | R H L      | P . I P       |     |
| CTCCCCAACT  | TATT       |             |            |               | 764 |
| L P N L     |            |             |            |               |     |
| S P T       | Y          |             |            |               |     |
| P Q L       | I          |             |            |               |     |

23 / 32

FIG 10

| 10         | 20         | 30         | 40         | 50          |     |
|------------|------------|------------|------------|-------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890  |     |
| TGTCGCTGT  | GCTCCTGATC | CAGCACAGGC | GCCCATTGCC | TCTCCCAATT  | 50  |
| C P L C    | S . S      | S T G      | A H C L    | S Q L       |     |
| V R C      | A P D P    | A Q A      | P I A      | S P N W     |     |
| S A V      | L L I      | Q H R R    | P L P      | L P I       |     |
| GGGCTAAAGG | CTTGCCATTG | TTCCTGCACA | GCTAAGTGCC | TGGGTTTCATC | 100 |
| G . R      | L A I V    | P A Q      | L S A      | W V H P     |     |
| A K G      | L P L      | F L H S    | . V P      | G F I       |     |
| G L K A    | C H C      | S C T      | A K C L    | G S S       |     |
| CTAATCGAGC | TGAACACTAG | TCACTGGGTT | CCACGGTTCT | CTTCCATGAC  | 150 |
| N R A      | E H .      | S L G S    | T V L      | F H D       |     |
| L I E L    | N T S      | H W V      | P R F S    | S M T       |     |
| . S S      | . T L V    | T G F      | H G S      | L P . P     |     |
| CCATGGCTTC | TAATAGAGCT | ATAACACTCA | CTGCATGGTC | CAAGATTCCA  | 200 |
| P W L L    | I E L      | . H S      | L H G P    | R F H       |     |
| H G F      | . . S Y    | N T H      | C M V      | Q D S I     |     |
| M A S      | N R A      | I T L T    | A W S      | K I P       |     |
| TTCCTTGGA  | TCCGTGAGAC | CAAGAACCCC | AGGTCAGAGA | ACACAAGGCT  | 250 |
| S L E      | S V R P    | R T P      | G Q R      | T Q G L     |     |
| P W N      | P . D      | Q E P Q    | V R E      | H K A       |     |
| F L G I    | R E T      | K N P      | R S E N    | T R L       |     |
| TGCCACCATG | TTGGAAGCAG | CCCACCACCA | TTTTGGAAGC | GGCCCCGCCAC | 300 |
| P P C      | W K Q      | P T T I    | L E A      | A R H       |     |
| C H H V    | G S S      | P P P      | F W K R    | P A T       |     |
| A T M      | L E A A    | H H H      | F G S      | G P P L     |     |
| TATCTTGGGA | GCTCTGGGAG | CAAGGACCCC | CAGGTAACAA | TTTGGTGACC  | 350 |
| Y L G S    | S G S      | K D P      | Q V T I    | W . P       |     |
| I L G      | A L G A    | R T P      | R . Q      | F G D H     |     |
| S W E      | L W E      | Q G P P    | G N N      | L V T       |     |
| ACGAAGGGAC | CTGAATCCGC | AACCATGAAG | GGATCTCCAA | AGCAATTGGA  | 400 |
| R R D      | L N P Q    | P . R      | D L Q      | S N W K     |     |
| E G T      | . I R      | N H E G    | I S K      | A I G       |     |
| T K G P    | E S A      | T M K      | G S P K    | Q L E       |     |

24 / 32  
**FIG 10 (suite)**

| 10         | 20         | 30          | 40          | 50          |     |
|------------|------------|-------------|-------------|-------------|-----|
| 1234567890 | 1234567890 | 1234567890  | 1234567890  | 1234567890  |     |
| AATGTTCTCT | CCAAGGCAAA | AATGCCCCCTA | AGATGTATTTC | TGGAGAATTG  | 450 |
| C S S      | Q G K      | N A P K     | M Y S       | G E L       |     |
| N V P P    | K A K      | M P L       | R C I L     | E N W       |     |
| M F L      | P R Q K    | C P .       | D V F       | W R I G     |     |
| GGACCAATCT | GACCTCAGA  | CAGTAAGAAA  | AAAAATGACT  | TATATTCTTC  | 500 |
| G P I .    | P S D      | S K K       | K N D L     | Y S S       |     |
| D Q S      | D P Q T    | V R K       | K M T       | Y I L L     |     |
| T N L      | T L R      | Q .         | E K         | K . L I F F |     |
| TGCAGTACCG | CCTGGCCACG | GATATCCTCT  | TCAAGGGGGA  | GAAACCTGGC  | 550 |
| A V P      | P G H G    | Y P L       | Q G G       | E T W P     |     |
| Q Y R      | L A T      | D I L F     | K G E       | K P G       |     |
| C S T A    | W P R      | I S S       | S R G R     | N L A       |     |
| CTCCTGAGGG | AAGTATAAAT | TATAACACCA  | TCTTACAGCT  | AGACCTGTTT  | 600 |
| P E G      | S I N      | Y N T I     | L Q L       | D L F       |     |
| L L R E    | V .        | I I T P     | S Y S .     | T C F       |     |
| S .        | G K Y K L  | . H H       | L T A       | R P V L     |     |
| TGTAGAAAAG | GAGGCAAATG | GAGTGAAGIG  | CCATATTIAC  | AAACTTTCTT  | 650 |
| C R K G    | G K W      | S E V       | P Y L Q     | T F F       |     |
| V E K      | E A N G    | V K C       | H I Y       | K L S F     |     |
| . K R      | R Q M      | E .         | S A         | I F T N F L |     |
| TTCATTAAAA | GACAACTCGC | AATTATGTAA  | ACAGIGTGAT  | TTGTGTCTTA  | 700 |
| S L K      | D N S Q    | L C K       | Q C D       | L C P T     |     |
| H .        | K T T R    | N Y V N     | S V I       | C V L       |     |
| F I K R    | Q L A      | I M .       | T V .       | F V S Y     |     |
| CAGGAAGCCC | TCAGATCTAC | CTCCCTACCC  | CGGCATCTCC  | CTGACTCCTT  | 750 |
| G S P      | Q I Y      | L P T P     | A S P       | . L L       |     |
| Q E A L    | R S T      | S L P       | R H L P     | D S F       |     |
| R K P      | S D L P    | P Y P       | G I S       | L T P S     |     |
| CCCCAACTAA | TAAGGACCCA | CTTCAGCCCA  | AACAGTCCAA  | AAGGACATAG  | 800 |
| P Q L I    | R T H      | F S P       | N S P K     | G H         |     |
| P N .      | . G P T    | S A Q       | T V Q       | K D I       |     |
| P T N      | K D P      | L Q P K     | Q S K       | R T .       |     |

25 / 32

FIG 11

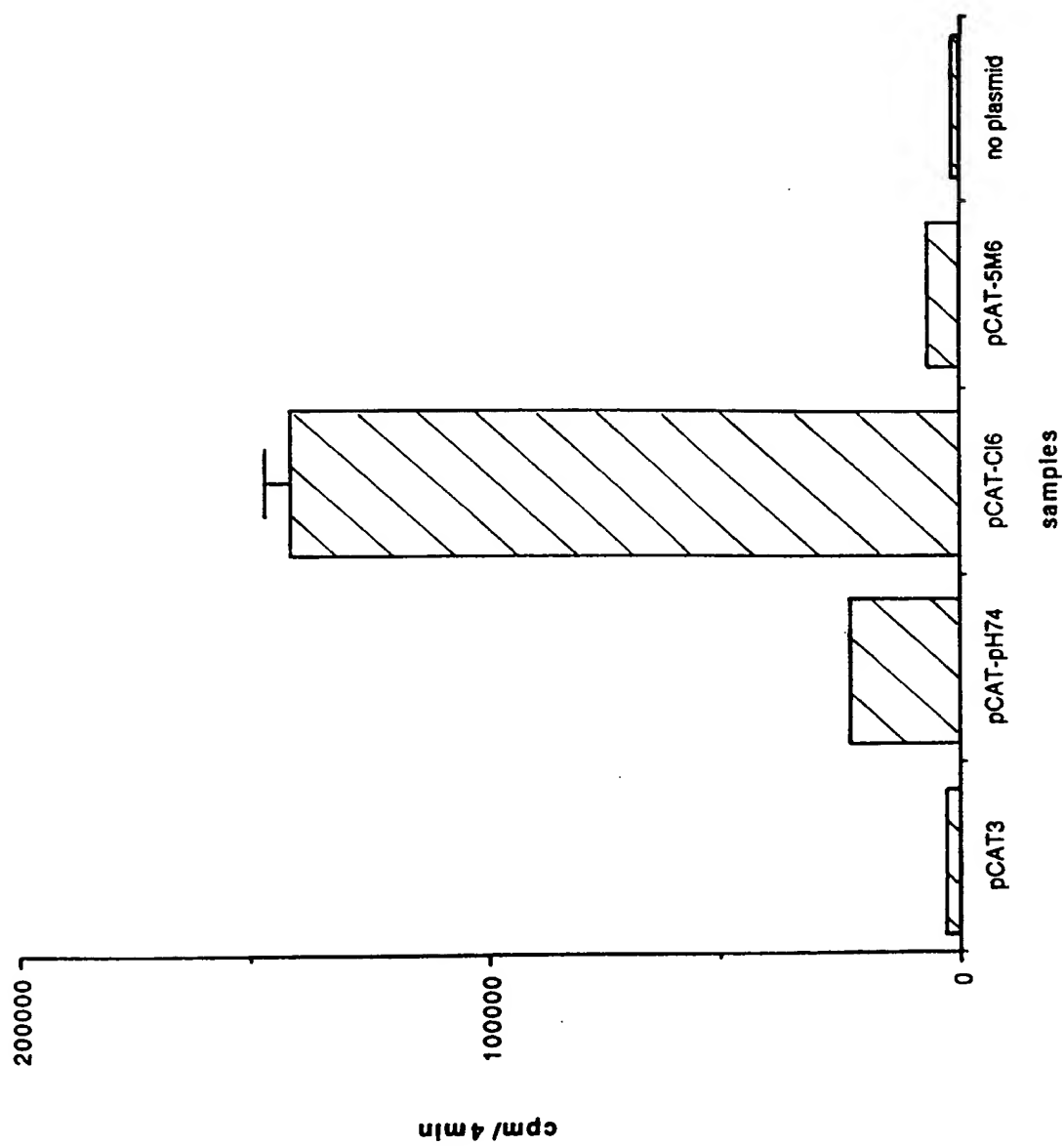
| 10         | 20         | 30         | 40          | 50         |     |
|------------|------------|------------|-------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890  | 1234567890 |     |
| GGCATTGATA | GCACCCATCA | GATGGCCAAA | TCATTATTTA  | CTGGACCAGG | 50  |
| G I D S    | T H Q      | M A K      | S L F T     | G P G      |     |
| A L I      | A P I R    | W P N      | H Y L       | L D Q A    |     |
| H . .      | H P S      | D G Q I    | I I Y       | W T R      |     |
| CCTTTTCAAA | ACTATCAAGC | AGATAGGGCC | CGTGAAGCAT  | GCCAAAGAAA | 100 |
| L F K      | T I K Q    | I G P      | V K H       | A K E I    |     |
| F S K      | L S S R    | . G P      | . S M       | P K K      |     |
| P F Q N    | Y Q A      | D R A      | R E A C     | Q R N      |     |
| TAATCCCCTG | CCTTATCGCC | ATGTTCTTTC | AGGAGAACAA  | AGAACAGGCC | 150 |
| I P C      | L I A      | M F L Q    | E N K       | E Q A      |     |
| . S P A    | L S P      | C S F      | R R T K     | N R P      |     |
| N P L      | P Y R H    | V P S      | G E Q       | R T G H    |     |
| ATTACCCAGG | GGAAGACTGG | CAACTAGATT | TTACCCACAT  | GGCCAAATGT | 200 |
| I T Q G    | K T G N    | . I L      | P T W       | P N V      |     |
| L P R      | G R L A    | T R F      | Y P H       | G Q M S    |     |
| Y P G      | E D W      | Q L D F    | T H M       | A K C      |     |
| CAGGGATTTC | AGCATCTACT | AGTCTGGGCA | GATACTTTCA  | CTGGTIGGGT | 250 |
| R D F      | S I Y .    | S G Q      | I L S       | L V G W    |     |
| G I S      | A S T      | S L G R    | Y F H       | W L G      |     |
| Q G F Q    | H L L      | V W A      | D T F T     | G W V      |     |
| GGAGTCTTCT | CCTTGTAGGA | CAGAAAAGAC | CCAAGAGGTA  | ATAAAGGCAC | 300 |
| S L L      | L V G      | Q K R P    | K R .       | . R H      |     |
| G V F S    | L . D      | R K D      | P R G N     | K G T      |     |
| E S S      | P C R T    | E K T      | Q E V       | I K A L    |     |
| TAATGAAATA | ATTCCCAGAT | TTGGACTTCC | CCCAGGATTIA | CAGGGTGACA | 350 |
| . . N N    | S Q I      | W T S      | P R I T     | G . Q      |     |
| N E I      | I P R F    | G L P      | P G L       | Q G D N    |     |
| M K .      | F P D      | L D F P    | Q D Y       | R V T      |     |

26 / 32  
**FIG 11 (suite)**

| 10         | 20         | 30         | 40         | 50         |     |
|------------|------------|------------|------------|------------|-----|
| 1234567890 | 1234567890 | 1234567890 | 1234567890 | 1234567890 |     |
| ATGGCCCCGC | TTTCAAGGCT | GCAGTAACCC | AGGGAGTATC | CCAGGIGTGA | 400 |
| W P R      | F Q G C    | S N P      | G S I      | P G V R    |     |
| G P A      | F K A      | A V T Q    | G V S      | Q V L      |     |
| M A P L    | S R L      | Q .        | P R E Y P  | R C .      |     |
| GGCATACAAT | ATCACTTACA | CTGTGCCTGG | AGGCCACAAT | CCTCCAGAAA | 450 |
| H T I      | S L T      | L C L E    | A T I      | L Q K      |     |
| G I Q Y    | H L H      | C A W      | R P Q S    | S R K      |     |
| A Y N      | I T Y T    | V P G      | G H N      | P P E K    |     |
| AGTCAAGAAA | ATGAATGAAA | CACTCAAAGA | TCTAAAAAAG | CTAACCCAAG | 500 |
| S Q E N    | E .        | N T Q R    | S K K A    | N P R      |     |
| V K K      | M N E T    | L K D      | L K K      | L T Q E    |     |
| S R K      | . M K      | H S K I    | . K S      | . P K      |     |
| AAACCCACAT | TGCATGACCT | GTTCGTGTGC | CTATAACCTT | ACTAAGAATC | 550 |
| N P H      | C M T C    | S V A      | Y N L      | T K N P    |     |
| T H I      | A .        | P V L L P  | I T L      | L R I      |     |
| K P T L    | H D L      | F C C      | L .        | P Y . E S  |     |
| CATAACTATC | CCCCAAAAAG | CAGGACTTAG | CCCATACGAG | ATGCTATATG | 600 |
| . L S      | P K K      | Q D L A    | H T R      | C Y M      |     |
| H N Y P    | P K S      | R T .      | P I R D    | A I W      |     |
| I T I      | P Q K A    | G L S      | P Y E      | M L Y G    |     |
| GATGGCCTTT | CCTAACCAAT | GACCTTGIGC | TTGACTGAGA | AATGGCCAAC | 650 |
| D G L S    | . P M      | T L C      | L T E K    | W P T      |     |
| M A F      | P N Q .    | P C A      | . L R      | N G Q L    |     |
| W P F      | L T N      | D L V L    | D .        | E M A N    |     |
| TTAGTTGCAG | ACATCACCTC | CTTAGCCAAA | TATCAACAAG | TTCTTAAAAC | 700 |
| . L Q      | T S P P    | . P N      | I N K      | F L K H    |     |
| S C R      | H H L      | L S Q I    | S T S      | S . N      |     |
| L V A D    | I T S      | L A K      | Y Q Q V    | L K T      |     |

27 / 32  
FIG 11 (suite)

| 10         | 20         | 30         | 40          | 50          |
|------------|------------|------------|-------------|-------------|
| 1234567890 | 1234567890 | 1234567890 | 1234567890  | 1234567890  |
| ATCACAGGGA | ACCTGTCCCC | GAGAGGAGGG | AAAGGAAC TA | TTCCACCC TG |
| H R E      | P V P      | E R R E    | R N Y       | S T L       |
| I T G N    | L S P      | R G G      | K G T I     | P P W       |
| S Q G      | T C P R    | E E G      | K E L       | F H P G     |
| 750        |            |            |             |             |
| GTGACATG   |            |            |             |             |
| 758        |            |            |             |             |
| V T        |            |            |             |             |
| . H        |            |            |             |             |
| D M        |            |            |             |             |





ATGGCCCTCC CTATACATAC TTTCCTCTTT ACTGCTCTCT TACCCCTTTT GGTCTCTACT GACCCCTCTC CATCTCTCTG TACACCACT AGCTCCCTCT  
 M A L P Y H T F L P T V L L P P F A L T A P P P C C C T T T B S S P Y  
 100 34  
 ACCAAGAGTT TCTATGAGA AGCGGCTTC CTGGAAATAT TGATGCCCA TCAATAGGA GTTATCTTAA GGGAACTCC ACCTTCACTG CCCACACCA  
 Q E F L R T R L P G N I D A P S Y R S L S K G N S T F T A H T H  
 200 67  
 TATGCCCGC AACTCTATA ACTCTCCAC TCTTTCATG CATGCAATA CTCTATATG GACAGGAAA ATGATTAATC CTATCTCTCT TGGAGACTT  
 H P R N C Y N S A T L C M H A N T H Y W T G K M I N P S C P G G L  
 300 100  
 GAGGCACTG TCTGTGGAC TTACTTACC CATACAGTA TGCTGTATG GGTGGAA TT CAAGTCAAG CAGAGAAA ACAAGTAAG GAGCAATCT  
 G A T V C W T Y P T H T S M S D G G G I Q G O A R E K Q V K E A I S  
 400 134  
 CCCAACTGAC CCGGGAACAT AGCACCCCTA CCCCTACAA AGCACTAGTT CTCTCAAAAC TACATGAAC CCTCGTACC CATATCTGOC TGGTGAAGCTT  
 Q L T R G H S T P S P Y K G L V L S K L H E T L R T H T R L V S L  
 500 167  
 ATTAAAGC ACCCTACTC GGTCTCAAG GGTCTAGCC CAAGCCCTA CTAACTGTG GATGTGCTC CCCTCACT TCAGCCATA CATTTCAATC  
 F N T T L T R L H E V S A Q N P T N C W H C L P L H F R P Y I S I  
 600 200  
 CTTCTCTG ACAGTGGAA CAATTCAGC ACAGAAATA ACAGCTTC GTTTTACTA GGACTCTTG TTCTCAATCT GGAAATAAC CATACCTCAA  
 P V P E Q W N N P S T E I N T T S V L V Q P L V S N L E I T H T S N  
 700 234  
 ACTCAGCTG TGTAAATTT AGCAATACTA TAGACACAC CACTCCCA TGCATCAGT GGTATACAC TCCACAGCA ATAGCTGOC TACCTCAGG  
 L T C V K P S N T I D T T S S Q C I R W V T P P T R I V C L P S G  
 800 267  
 AATATTTT CTCTGTGTA CTTCAAGCTA TCAITTTTG AAGCTCTT CAGAACTAT GTCTCTCTC TCAITTTTAG TGCCTCTAT GACCACTAC  
 I F P V C G T S A Y H C L N G S S E S M C F L S F L V P P M T I Y  
 900 300  
 ACTGAACAG ATTATACAA TCAITGTGTA CTTAAGCCC ACACAAAG AGTACCAAT CTTCCTTTG TTATCAGAC AGGAGCTTA GGCAGCTAG  
 T E Q D L Y N H V V P K P H N K R V P I L P F V I R A G V L G R L G  
 1000 334  
 GTACTGGAT TGGGATATC ACAACCTTA CTCAGTTCTA CTCAACTA TCTCAAGAA TAAATGTGA CATGGAACG GTCACTGACT CCTGTGTCAC  
 T G I G S I T T S T Q F Y Y K L S Q E I N G D H E Q V T D S L V T  
 1100 367  
 CTTGCAAGT CAATTAAC CCGTACAGC AGTAGTCTT CAATTCGAA GAGCTTAGA CTTGCTAAC GCCAAAGAG GGGCACTG TTTATTTTA  
 L Q D Q L N S L A A V V L Q N R R A L D L L T A K R G G T C L F L  
 1200 400  
 GGAGAAGAC GCTGTATTA TGTAAATCA TCACTGAGAA AGTTAAGAA ATTGAGATC GAATACAA TGAGGAGAG GAGCTTCAA  
 G E R C Y Y V N S R I V T E K V K E I R D R I Q C R A E E L Q N  
 1300 434  
 AACCGAAG CTGGGCTC CTCAGCAAT GATGCCCTG GGTCTCTCC TTCTTAGAC CTCAGAGC TCTAATATG TTACTCTCT TTGGACCTG  
 T E R W G L L S Q W M P W V L P F L G P L A A L I L L L L P G P C  
 1400 467  
 TATCTTTAAG CTCCTGTGA AGTTGTCTC TTCCAGATT TTCCAGATT AGCTGATA AGCTACAGT GGTCTTACAA ATGGAACCCC AGATGGAGTC CATGACTAAG  
 I F N L L V K F V S S R I E A V K L Q M V L Q H E P Q H E S H T K  
 1500 500  
 ATCCAGCTG GACCCCTGGA CCGGCTGCT AGCCATGCT CCGATGTAA TGACATTGAA GGCACCTC CCGAGCAAT CTCAACTGCA CAACCTTAC  
 I H R G P L D R P A S P C S D V N D I E G T P P E I S T A Q P L L  
 1600 534  
 TATGCCCAA TTCAGGGA AGCAGTTAGA GGTATCA GGCACCTCC CCAACAGCAC TTGGTTTTC CTGTTGAGAG GGGGACTGA GAGACAGGAC  
 C P N S A G S  
 1700 542  
 TAGTGGATT TCTTAGGCA ACAGAAATC CTTAGCCTA GCTGGAGG TGACTCATC CACCTTAAA CATGGGCTT GCAACTTAGC TCACACCGA  
 1800  
 CCAATCAG AGCTCACTAA AATGCTAAT AGGCAAAAT AGGAGTAAA GAATAGCA ATCATCTATT GCTCAGAGC ACAGCGGAG GACAAAGAT  
 1900  
 CGGATATA ACCCAGGAT TCGAGCGGC AACGGCACT CTTTGGGT CCGCTCTCT TGTATGGCG CTCTGTTTC ACTCTATTTC ACTCTATTAA  
 2000  
 ATTTCGAC TGAATAAA AAAAAAAAAA  
 2030  
 Cap site  
 Poly A signal

FIG13

FIG 14

|          |               |               |               |               |               |             |            |            |            |                         |      |
|----------|---------------|---------------|---------------|---------------|---------------|-------------|------------|------------|------------|-------------------------|------|
| <b>R</b> | CAGCAACCCC    | CTTTGGGTCC    | CCTCCCATG     | TATGGAGCT     | CTGTTTTCAC    | TCTATTTCAC  | TCTATTAAAT | CATGCAACTG | CACTCTTCG  | GTCCGTGTTT              | 100  |
|          | TTTATGGCTC    | AAGCTGAGCT    | TTTGTTTGCC    | ATCCACCACT    | GCTGTTTGCC    | ACCGTCCAG   | ACCGCTGCT  | GACITCCATC | CCTTTGGATC | CAGCAGATG               | 200  |
|          | TCCGCTGTGC    | TCTGTATCCA    | GCACAGGCG     | CCATTGCGTC    | TCCCAATTGG    | GCTAAAGCT   | TGSCATTGTT | CCTGCACAGC | TAAGTGCCTG | GGTTTCATCT              | 300  |
|          | AATCGAGCTG    | AACACTAGTC    | ACTGGGTCC     | ACGGTCTCT     | TCCATGACCC    | ATGGCTTCTA  | ATAGAGCTAT | AACACTCACT | GCATGTCCA  | AGATTCCATT              | 400  |
|          | CCTTGGAAATC   | CGTGAGACCA    | AGAACCCCG     | GTCAGAGAAC    | ACAAGGCTTG    | CCACCATGTT  | GGAAGCAGCC | CACCACCAAT | TTGGAAGCAG | CCCGCCACTA              | 500  |
|          | TCTTGGAGC     | TCTGGAGCA     | AGGACCCCG     | GTAACAATTT    | GGTGACCACT    | AAGGGAAGCTG | AATCCGCAAC | CATGAAGGGA | TCTCCAAGC  | <sup>gag</sup> ATGGGAAC | 600  |
|          | GTTCCTCCCG    | AGGCAAAAT     | GGCCCTAGAA    | CGTATCTGG     | AGAATTGGGA    | CCAATGTGAC  | ACTCAGAGC  | TAAGAAGAA  | ACGATTATA  | TTCTTCTGCA              | 700  |
|          | V P P E A K M | P L E R I L E | N W D Q C D   | T Q T L R K K | R F I F C S   |             |            |            |            |                         | 37   |
|          | GTACCGCTG     | GCCACAATAT    | CCTCTTCAAG    | GGAGAGAAC     | CTGGCTTCT     | GAGGGAAGTA  | TTAATTATAA | CATCATCTTA | CAGCTAGACC | TCTTCTGTAG              | 800  |
|          | T A W P Q Y   | P L Q G R E T | W L P E G S I | N Y N I I L   | Q L D L F C R |             |            |            |            |                         | 70   |
|          | AAAGGAGGC     | AAATGGAGTG    | AAGTGCCATA    | TGTGCAAACT    | TTCTTTTCAT    | TAAGAGACAA  | CTCACAATTA | TGTAAAAAGT | GTGGTTTATG | CCTACAGGA               | 900  |
|          | K E G K W S   | E V P Y V Q T | F F S L R D N | S Q L C K K C | G L C P T G   |             |            |            |            |                         | 103  |
|          | AGCCCTCAGA    | GTCCACCTCC    | CTACCCCGAG    | GTCCCTCCC     | CGACTCCTTC    | CTCAACTAAT  | AAGGACCCC  | CTTTAACCCA | AAGGTCCAA  | AAGGAGATAG              | 1000 |
|          | S P Q S P P   | P Y P S V P S | T P S S T N   | K D P P L T Q | T V Q K E I D |             |            |            |            |                         | 137  |
|          | ACAAAGGGT     | AAACAATGAA    | CCAAAGAGTG    | CCAATATTC     | CCGATTATGC    | CCCCTCCAG   | CAGTCAGAGG | AGGAGAATTC | GGCCGAGCCA | GAGTGCCTGT              | 1100 |
|          | K G V N N E   | P K S A N I P | R L C P L Q A | V R G G E F   | G P A R V P V |             |            |            |            |                         | 170  |
|          | ACCTTTTCT     | CTCTCAGACT    | TAAAGCAAT     | TAAATAGAC     | CTAGGTAAT     | TCTCAGATAA  | CCCTGACGGC | TATATTGATG | TTTTACAAGG | GTTAGGACAA              | 1200 |
|          | P F S L S D   | L K Q I K I D | L G K F S D   | N P D G Y I D | V L Q G L G Q |             |            |            |            |                         | 203  |
|          | TCTTTGATC     | TGACATGAG     | AGATATAATG    | TTACTACTAA    | ATCAGACACT    | AACCCCAAT   | GAGAGAAGTG | CGCTGTATAC | TGCAGCCGA  | GAGTTTGGCG              | 1300 |
|          | S F D L T W   | R D I M L L L | N Q T L T P N | E R S A A V T | A A R E F G D |             |            |            |            |                         | 237  |
|          | ATCTTGGTA     | TCTCAGTCAG    | GCCACAATA     | GGATGACAC     | AGAGAAAGA     | ACAACTCCA   | CAGGCCAGCA | GGCAGTCCC  | AGTGTAGACC | CTCATTTGGA              | 1400 |
|          | L W Y L S Q   | A N N R M T T | E E R T T P T | G Q Q A V P   | S V D P H W D |             |            |            |            |                         | 270  |
|          | CACAGAATCA    | GAACATGAG     | ATTGGTGCCA    | CAACATTTG     | CTAATTGGC     | TGCTAGAAG   | ACTGAGAAA  | AGCCTATGAA | TTACTCAATG |                         | 1500 |
|          | T E S E H G   | D W C H K H L | L T C V L E G | L R K T R K K | P M N Y S M   |             |            |            |            |                         | 303  |
|          | ATGTCCACTA    | TAACACAGG     | AAAGGAAGA     | AATCTTACTG    | CTTTTCTGGA    | CAGACTAAGG  | GAGGCAATGA | GGAGCATATC | CTCCCTGTCA | CCTGACTCTA              | 1600 |
|          | M S T I T Q   | G K E E N L T | A F L D R E A | L R K K H T   |               |             |            |            |            |                         | 337  |
|          | TTGAAGGCCA    | ACTAATCTTA    | AAGGATAAGT    | TTATCACTCA    | GTCACTGCA     | GACATTAGAA  | AAAACTTCA  | AAAGTCCGTC | TTAGGCTCGG | AAACAAACTT              | 1700 |
|          | E G Q L I L   | K D K F I T Q | S A A D I R K | K L Q K S V   | L G S E Q N L |             |            |            |            |                         | 370  |
|          | AGAAACCCCTA   | TTGAATCTGG    | CAACCTCGGT    | TTTTTATAT     | AGAGATCAGG    | AGGAGAGGC   | AGATGGGAC  | AAATGGGATA | AAAAAAAAG  | GGCCACCGCT              | 1800 |
|          | E T L L N L   | A T S V F Y N | R D Q E E Q A | E W D K W D   | K K R A T A   |             |            |            |            |                         | 403  |
|          | TTAGTCATGG    | CCCTCAGGCA    | ACGGGACTTT    | GGAGGCTCTG    | GAAAGGGAA     | AAGCTGGCA   | AATAGGAAGC | CTAATAGGGC | TTGCTTCCAG | TGCGTCTTAC              | 1900 |
|          | L V M A L R   | Q A D F G G S | G K G K S W   | A N R K P N R | A C F Q G L Q |             |            |            |            |                         | 437  |
|          | AAGGACACTT    | TAAATAAGAT    | TGTCCAATA     | GAATAAGCC     | GGCCCTTGT     | CCATGCCCTT  | TAGTCAAGG  | GAATCACTGG | AAGGCCCACT | GGCCGAGGGG              | 2000 |
|          | G H F K K D   | C P N R N K P | P P P C/R P   | C P L R Q G N | / H W K A H C | P R G       |            |            |            |                         | 470  |
|          | ATCAAGATAC    | TCTGAGTCAG    | AAGCCATTAA    | CCAGATGATC    | CAGCAGCAGG    | ACTGA       |            |            |            |                         | 2055 |
|          | S R Y S E S   | E A I N Q M I | Q Q Q D       |               |               |             |            |            |            |                         | 487  |

FIG 15

100 GGACCCGTAG TATGGGGTAA TCCCCTCCGG GAACCCNAGC CCCAGTACTC AGAAGAAGAA ATAGAATGGG GAACCTCAGC AGGACATGGT TTCTCCCT  
 34 G P V V W G N P L R E T K P Q Y S E E E I E W G T S R G H G F L P S  
 200 CAGGATGGCT AGCCACTGAA GAAGGAJAAA TACTTTTGCT GGCAGCTAAC CAATGGAAAT TACTTAAJAC CTTTCAGCAA ACCTTCCACT TAGGCAITGA  
 67 G W L A T E E G K I L L L A A N Q W K L L K T L Q Q T F L L G I D  
 300 TAGCACCCAT CAGATAGCCA AATCATTTAT TACTGGACCA GGCCTTTTCA AAATATATCA GCAGATAGTC AGGGCCTGTG AAGTGTGCCA AAGAATTAAT  
 100 S T M Q I A K S L F T G P G L F K T I K Q I V R A E V E Q R N M  
 400 CCCCTGCCCT ATGCCAAGC TCCTTCAGGA GAACAAAGAA CAGGCAATTA CCCAAGAGAA GACTGGCAAC TAGATTTTAT CCACATGCCA AAATCAGAG  
 134 P L P Y R Q A P S G E Q R T G N Y P R E D W Q L D F I H M P K S Q G  
 500 GATTTCAGTG TCTACTAGTC TGGGTAGATA CTTTCACTGG TTGGGCAGAG GCCTTCCCT GTAGACAGA AAAGTTCCAA GAGGTAAATA AGGCACTAGT  
 167 F Q C L L V W V D T F T G W A E A P P C R T E K F Q E V I K A L V  
 600 TCATGAAGTA ATTCCAGAT TCGACTTCC CTGAGGCTTA CAGAGTGACA ATGGTCTGTC TTTCAGGCC ACAGTAACCC AGGAGTATC CCAGGCGTTA  
 200 H E V I P R F G L P . G L Q S D N G P A F K A T V T Q G V S Q A L  
 700 GGTATAGAAT ATCACTTACA CTGCACCTAG AGGCCAAT CTTCAGGGA GGTTCAGAA ATGAJACAC TCAAGACAA TCTTAACAG CTAAACCCAG  
 234 G I E Y H L H C T . R P Q S S G K V E K M K T L K R H L N K L T Q E  
 800 AAACCCACCT CGCATGGTCT GCTCTGTGT CTATAGCCTT ACTAAGATC CAUACTCTC CCCAAAGGC AGGACTTAGC CCATACAGAA TGTGTATGG  
 267 T H L A W S A L L S I A L L R I Q N S P Q K A G L S P Y R M L Y G  
 900 ACGGTCTCTC CTAAACCAATG ACCTTCTGCT TGACCAAGAG ATGGCCAACT TAGTTCAGA CATCACCCTC TTAGCCAAAT ATCAACAAAT TCTTAACA  
 300 R S F L T N D L L L D Q E M A N L V A D I T S L A K Y Q Q V L K T  
 1000 TTACAAAGAG CCTGTCCCG AGAGGAGGA AAGAATAT TCCACCTGG TGTATGGTA TAGTCAAGT CCTTCCCTC TAATTCCTCA TCCTTAGACA  
 334 L Q G A C P R E E G K E I F H P G V M V L V K S L P S M S P S L D T  
 1100 CATCTGGG AGGACCTTAC CCAGTCATTT TATCTATCCC AACTGCGTT AAGTGGCTG GAGTGGAGTC TTGATACAT CACACTCGAA TCAAGCCCTG  
 367 S W G G P Y P V I L S I P T A V K V A G V E S W I H H T R I K P M  
 1197 GATACTGCG AGGAAACCG AAATCCAGG GGACACCGT AGCTATTTCT TTGAACCTCT AGAGATCTG TCCTGTCTT TCAGCAACA ACCGTGA  
 398 I L P K E P E N P G D N A S Y F F E P L E D L C L L F K Q Q P

FIG 16

100 GAGATAGCA GGTATGTTG CTTGGCAGCA GTATGGAGG AACACAGAA GTACAGAAA GAAAGAGAA GAGGAGAAA  
 E.N.S.S.I.S.W.L.A.E.V.G.K.D.S.K.K.R.K.K.E.S.Q.R.K.K.R.E.E.E.T  
 200 CAACAGCAA CTTGAGGAA GAAAGAGTA GTAAAGAAA AACAGATAC CTTATCTTT TAAGAGTTG GCTAATTTT TGTCTACTA CCAAGGCAAT  
 K.K.N.L.K.R.E.R.S.S.K.E.K.T.V.Y.P.I.P.L.K.A.R.V.N.F.C.L.P.S.Q.G.I  
 300 XTCTTCTTA TGTGCAAT CACCTTAT CTTCTTCCC ACTACTTCA CAGGACGAG AACCTTGTG TTTCTATTC CCAACATTA CATTECCCA  
 F.F.L.C.G.T.S.T.Y.I.C.L.P.T.N.W.T.G.T.R.T.L.V.F.L.S.P.N.I.N.I.A.P  
 400 GGAATGAGA CTTATTTGT ACCGTCAAA GCTAAGTTC GTAGAGTGG ACCACAAA CTAATATCC TATTTAGGG GTAGAGATG CTTACTTCTTA  
 G.N.Q.T.L.L.V.P.V.K.A.K.V.R.Q.C.R.A.I.Q.L.I.S.L.F.I.G.L.G.M.A.T.A.T  
 500 CAGAGATGG AATAGCGTT TPTCTACT CATTACTTA CTTCAATGA ATTCTTGA CAGTTTGA GAAATATGA AATATATCT  
 G.T.G.I.A.G.L.S.T.S.L.S.Y.Y.H.T.L.S.K.N.F.S.D.S.L.Q.E.I.M.K.S.I.L  
 600 TACTTTCAA TCCCATNG ACTCTTGG ACATATCT CTTCAATC GCGAGGCC ACACCTCTC ACTGTCGA AAGAGGCT CTTACCTTC  
 T.L.Q.S.Q.L.D.S.L.A.A.M.T.L.Q.N.R.R.G.P.H.L.L.T.A.E.K.G.G.L.C.T.F  
 700 TTAGGAGAG AATGTTT TTTCACTAC CAGTACGA TAGTACGA TCCACTGG CATTTAGG AAGGCTTC TGTATAGA CAATGCTTT  
 L.G.E.C.C.F.Y.T.N.O.S.G.I.V.R.D.A.T.W.H.L.Q.E.R.A.S.D.I.R.Q.C.L.S  
 800 CAATCTTA TCCATCTC TGAATGG CAGATGGT TCTTCTAT CTTAGTCA TGGAGGCT CTTCTTGA CTTACCTTC GGCCTGTAT  
 N.S.Y.T.N.L.W.S.W.A.T.W.L.L.P.F.L.G.P.M.A.A.I.L.L.L.L.T.F.G.P.C.I  
 900 TTTTAACTT CTGTCAAT TTTTCTTC TAGATGAA CCAATCAC TACAGTGT CTTCAATG GAACCAAA TGTCTTCA TACATCTC  
 F.K.L.L.V.K.F.V.S.S.R.I.E.A.I.K.L.Q.M.V.L.Q.M.E.P.Q.M.S.S.T.N.N.F  
 1000 TACCAATAC CTTGAGAG ATTCACTTC ACTTCACTA CTTTCAAT TCCCTTGG AAGACATAC AACTGAGG CTTCTTCTT GGCCTATCC  
 Y.Q.G.P.L.E.R.S.T.G.T.S.T.S.L.E.I.P.L.W.K.T.L.Q.L.Q.G.P.F.F.A.P.I.Q  
 1100 ACGAGAGT AGCTAGAG GTCATGGC AATTTCCA CAGATTTG GGTCTTGT TTAGGGGG GATTCAGG TACACTT CTTGGAGC  
 Q.E.V.A.R.A.V.I.G.Q.I.P.N.S.S.W.G.V.L.F.R.G.G.I.E.E.A.C.W.Q.P  
 1200 TCACAGCTT CTTGATCT CAGTCTTC TCCACTCT GGCCTTCTT CAGTCTCT TACGCTTC ACTGACTCT CAGGCTCT  
 H.S.P.R.W.I.S.V.P.P.Q.P.W.C.P.L.W.P.C.L.R.S.P.S.A.C.H.C.T.V.G.A.S  
 1300 TTTCTGGTG GTCAGGCT CATTGCTG CAGGAGTA TGGAGAGA GATGAGAG CTTGAGAG CTTGAGAG CTTTGGGG  
 F.W.A.G.Q.G.R.S.Q.L.P.Q.L.A.G.R.Y.G.G.R.D.A.G.G.N.Q.G.C.A.W.R.L.R.A  
 1400 CAGGCTTG TTTGAGTG GGTGGCTC GCGGCTCC ACATGAGC AATGAGGC TTGAGCTG GCGAGAG ATGCTTCTT CAACTTCTC  
 S.M.S.S.R.W.A.W.A.R.R.A.P.H.S.G.S.E.G.L.S.T.W.A.R.Q.M.L.C.S.T.S.S  
 1500 GCTGGCTT AGTCTCTC CTTTCTTC AGCTTACG GATCTGAG CTTGATAG CTTAGTCC CACCTTCTT GTGCTTCTT GAGAGGCA  
 L.G.L.S.C.L.P.R.G.A.G.L.R.E.H.A.A.C.P.C.L.S.P.P.P.R.R.G.F.L.H.S.P  
 1600 AGCTTGG AGAAGCTA CTTTATTC AGGTGCTA GTTCATGA CCAACAGG GTTGAAT GCGGCTAC AGGCGGAT TCGAGGAG  
 S.F.P.D.K.H.H.P.L.S.T.V.P.S.P.I.N.H.P.R.V.E.E.C.G.H.T.A.R.D.W.Q.A.V  
 1700 TTTCACTGC CTTCTTGT CCGATCAC TCCATGAG CAGTGGCT CTTGATAG GTGGCTCT CAGATCTT TATGCTTCTT TATGCTTCTT  
 P.L.A.A.L.V.R.D.P.L.R.E.A.S.W.A.P.E.S.G.G.D.L.E.N.L.Y.V.L.R.D.C  
 TAAATACAC AATCAGAC  
 K.Y.T.N.Q.H

1719

## LISTE DE SEQUENCES

## (2) INFORMATIONS POUR LA SEQ ID NO: 68:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- 5 (A) LONGUEUR: 34 paires de bases  
(B) TYPE: nucléotide  
(C) NOMBRE DE BRINS: simple  
(D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADNc

## 10 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 68:

GACTCGCTGC AGATCGATTT TTTTTTTTTT TTTT

34

## (2) INFORMATIONS POUR LA SEQ ID NO: 69:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- 15 (A) LONGUEUR: 30 paires de bases  
(B) TYPE: nucléotide  
(C) NOMBRE DE BRINS: simple  
(D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADNc

## 20 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 69:

GCCATCAAGC CACCCAAGAA CTCTTAAGT

30

## (2) INFORMATIONS POUR LA SEQ ID NO: 70:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- 25 (A) LONGUEUR: 30 paires de bases  
(B) TYPE: nucléotide  
(C) NOMBRE DE BRINS: simple  
(D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADNc

## 30 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 70:

CCAATAGCCA GACCATTATA TACACTAATT

30

## (2) INFORMATIONS POUR LA SEQ ID NO: 112:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- 35 (A) LONGUEUR: 310 paires de bases  
(B) TYPE: nucléotide

2

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 112:

```

5 GCTTATAGAA GGACCCCTAG TATGGGGTAA TCCCCTCTGG GAAACCAAGC CCCAGTACTC 60
 AGCAGGAAAA ATAGAATAGG AAACCTCACA AGGACATACT TTCCTCCCCT CCAGATGGCT 120
 AGCCACTGAG GAAGGAAAAA TACTTTCACC TGCAGCTAAC CAACAGAAAT TACTTAAAAC 180
 CCTTCACCAA ACCTTCCACT TAGGCATTGA TAGCACCCAT CAGATGGCCA AATTATTATT 240
 TACTGGACCA GGCCTTTTCA AAACATATCAA GAAGATAGTC AGGGGCTGTG AAGTGTGCCA 300
10 AAGAAATAAT 310

```

(2) INFORMATIONS POUR LA SEQ ID NO: 113:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 103 acides aminés

15 (B) TYPE: acide aminé

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: peptide

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 113:

```

20 Leu Ile Glu Gly Pro Leu Val Trp Gly Asn Pro Leu Trp Glu Thr Lys
 1 5 10 15
 Pro Gln Tyr Ser Ala Gly Lys Ile Glu Xaa Glu Thr Ser Gln Gly His
 20 25 30
 Thr Phe Leu Pro Ser Arg Trp Leu Ala Thr Glu Glu Gly Lys Ile Leu
25 35 40 45
 Ser Pro Ala Ala Asn Gln Gln Lys Leu Leu Lys Thr Leu His Gln Thr
 50 55 60
 Phe His Leu Gly Ile Asp Ser Thr His Gln Met Ala Lys Leu Leu Phe
 65 70 75 80
30 Thr Gly Pro Gly Leu Phe Lys Thr Ile Lys Lys Ile Val Arg Gly Cys
 85 90 95
 Glu Val Cys Gln Arg Asn Asn
 100

```

35 (2) INFORMATIONS POUR LA SEQ ID NO: 114:

(i) CARACTERISTIQUES DE LA SEQUENCE:

3

(A) LONGUEUR: 635 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

5 (ii) TYPE DE MOLECULE: ADNc

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 114:

```

CCCTGTATCT TTAACCTCCT TGTTAAGTTT GTCTCTTCCA GAATCAAAAC TGTA AAACTA 60
CAAATTGTTT TTCAAATGGA GCACCAGATG GAGTCCATGA CTAAGATCCA CCGTGGACCC 120
CTGGACCGGC CTGCTAGCCC ATGCTCCGAT GTTAATGACA TTGAAGGCAC CCCTCCCGAG 180
10 GAAATCTCAA CTGCACAACC CCTACTATGC CCCAATTCAG CGGGAAGCAG TTAGAGCGGT 240
CATCAGCCAA CCTCCCCAAC AGCACTTGGG TTTTCCTGTT GAGAGGGGGG ACTGAGAGAC 300
AGGACTAGCT GGATTCCTA GGCCAACGAA GAATCCCTAA GCCTAGCTGG GAAGGTGACT 360
GCATCCACCT CTAAACATGG GGCTTGCAAC TTAGCTCACA CCCGACCAAT CAGAGAGCTC 420
ACTAAAATGC TAATTAGGCA AAAATAGGAG GTAAAGAAAT AGCCAATCAT CTATTGCCTG 480
15 AGAGCACAGC GGGAGGGACA AGGATCGGGA TATAAACCCA GGCATTGAG CCGGCAACGG 540
CAACCCCTT TGGGTCCCT CCCTTTGTAT GGGCGCTCTG TTTTACTCT ATTTACTCT 600
ATTAAATCTT GCAACTGAAA AAAAAAAAAA AAAAA 635

```

(2) INFORMATIONS POUR LA SEQ ID NO: 115:

20 (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 77 acides aminés

(B) TYPE: acide aminé

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

25 (ii) TYPE DE MOLECULE: peptide

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 115:

```

Pro Cys Ile Phe Asn Leu Leu Val Lys Phe Val Ser Ser Arg Ile Lys
1 5 10 15
Thr Val Lys Leu Gln Ile Val Leu Gln Met Glu His Gln Met Glu Ser
30 20 25 30
Met Thr Lys Ile His Arg Gly Pro Leu Asp Arg Pro Ala Ser Pro Cys
35 40 45
Ser Asp Val Asn Asp Ile Glu Gly Thr Pro Pro Glu Glu Ile Ser Thr
50 55 60
35 Ala Gln Pro Leu Leu Cys Pro Asn Ser Ala Gly Ser Ser
65 70 75

```

## (2) INFORMATIONS POUR LA SEQ ID NO: 116:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- 5 (A) LONGUEUR: 32 paires de bases  
 (B) TYPE: nucléotide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADNc

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 116:

10 TGGGGTTCCA TTTGTAAGAC CATCTGTAGC TT 32

## (2) INFORMATIONS POUR LA SEQ ID NO: 117:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- 15 (A) LONGUEUR: 1481 paires de bases  
 (B) TYPE: nucléotide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADNc

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 117:

20 ATGGCCCTCC CTTATCATAC TTTTCTCTTT ACTGTTCTCT TACCCCTTT CGCTCTCACT 60  
 GCACCCCTC CATGCTGCTG TACAACCAGT AGCTCCCCTT ACCAAGAGTT TCTATGAAGA 120  
 ACGCGGCTTC CTGGAAATAT TGATGCCCCA TCATATAGGA GTTTATCTAA GGGAAACTCC 180  
 ACCTTCACTG CCCACACCCA TATGCCCCGC AACTGCTATA ACTCTGCCAC TCTTTCATG 240  
 CATGCAAATA CTCATTATTG GACAGGGAAA ATGATTAATC CTAGTTGTCC TGGAGGACTT 300  
 25 GGAGCCACTG TCTGTTGGAC TTA CTTCACC CATAACAGTA TGTCTGATGG GGGTGAATT 360  
 CAAGGTCAGG CAAGAGAAAA ACAAGTAAAG GAAGCAATCT CCCAACTGAC CCGGGGACAT 420  
 AGCACCCCTA GCCCCTACAA AGGACTAGTT CTCTCAAAAC TACATGAAAC CCTCCGTACC 480  
 CATACTCGCC TGGTGAGCCT ATTTAATACC ACCCTCACTC GGCTCCATGA GGTCTCAGCC 540  
 CAAAACCCTA CTA ACTGTTG GATGTGCCTC CCCCTGCACT TCAGGCCATA CATTTCAATC 600  
 30 CCTGTTCCCTG AACAATGGAA CAACTTCAGC ACAGAAATAA ACACCACTTC CGTTTTAGTA 660  
 GGACCTCTTG TTTCCAATCT GGAAATAACC CATACTCAA ACCTCACCTG TGTAAAATTT 720  
 AGCAATACTA TAGACACAAC CAGCTCCCAA TGCATCAGGT GGGTAACACC TCCCACACGA 780  
 ATAGTCTGCC TACCCTCAGG AATATTTTTT GTCTGTGGTA CCTCAGCCTA TCATTGTTTG 840  
 AATGGCTCTT CAGAATCTAT GTGCTTCCTC TCATTCTTAG TGCCCCCTAT GACCATCTAC 900  
 35 ACTGAACAAG ATTTATACAA TCATGTCGTA CCTAAGCCCC ACAACAAAAG AGTACCCATT 960  
 CTTCCCTTTTG TTATCAGAGC AGGAGTGCTA GGCAGACTAG GTACTGGCAT TGGCAGTATC 1020



ACAACCTCTA CTCAGTTCTA CTACAACTA TCTCAAGAAA TAAATGGTGA CATGGAACAG 1080  
 GTCAGTACT CCCTGGTCAC CTTGCAAGAT CAACTTAACT CCCTAGCAGC AGTAGTCCTT 1140  
 CAAAATCGAA GAGCTTTAGA CTTGCTAACC GCCAAAAGAG GGGGAACCTG TTTATTTTTA 1200  
 GGAGAAGAAC GCTGTTATTA TGTTAATCAA TCCAGAATTG TCACTGAGAA AGTTAAAGAA 1260  
 5 ATTCGAGATC GAATACAATG TAGAGCAGAG GAGCTTCAAA ACACCGAACG CTGGGGCCTC 1320  
 CTCAGCCAAT GGATGCCCTG GGTTCCTCCC TTCTTAGGAC CTCTAGCAGC TCTAATATTG 1380  
 TTACTCCTCT TTGGACCCTG TATCTTTAAC CTCCTTGTTA AGTTTGTCTC TTCCAGAATT 1440  
 GAAGCTGTAA AGCTACAGAT GGTCTTACAA ATGGAACCCC A 1481

10 (2) INFORMATIONS POUR LA SEQ ID NO: 118:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 493 acides aminés

(B) TYPE: acide aminé

(C) NOMBRE DE BRINS: simple

15 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: peptide

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 118:

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | Met | Ala | Leu | Pro | Tyr | His | Thr | Phe | Leu | Phe | Thr | Val | Leu | Leu | Pro | Pro |
|    | 1   |     |     |     | 5   |     |     |     | 10  |     |     |     |     | 15  |     |     |
| 20 | Phe | Ala | Leu | Thr | Ala | Pro | Pro | Pro | Cys | Cys | Cys | Thr | Thr | Ser | Ser | Ser |
|    |     |     |     | 20  |     |     |     | 25  |     |     |     |     | 30  |     |     |     |
|    | Pro | Tyr | Gln | Glu | Phe | Leu | Xaa | Arg | Thr | Arg | Leu | Pro | Gly | Asn | Ile | Asp |
|    |     | 35  |     |     |     |     | 40  |     |     |     | 45  |     |     |     |     |     |
|    | Ala | Pro | Ser | Tyr | Arg | Ser | Leu | Ser | Lys | Gly | Asn | Ser | Thr | Phe | Thr | Ala |
| 25 |     | 50  |     |     |     | 55  |     |     | 60  |     |     |     |     |     |     |     |
|    | His | Thr | His | Met | Pro | Arg | Asn | Cys | Tyr | Asn | Ser | Ala | Thr | Leu | Cys | Met |
|    | 65  |     |     |     | 70  |     |     |     | 75  |     |     |     |     | 80  |     |     |
|    | His | Ala | Asn | Thr | His | Tyr | Trp | Thr | Gly | Lys | Met | Ile | Asn | Pro | Ser | Cys |
|    |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |     |     |
| 30 | Pro | Gly | Gly | Leu | Gly | Ala | Thr | Val | Cys | Trp | Thr | Tyr | Phe | Thr | His | Thr |
|    |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |     |
|    | Ser | Met | Ser | Asp | Gly | Gly | Gly | Ile | Gln | Gly | Gln | Ala | Arg | Glu | Lys | Gln |
|    |     |     | 115 |     |     |     |     | 120 |     |     |     | 125 |     |     |     |     |
|    | Val | Lys | Glu | Ala | Ile | Ser | Gln | Leu | Thr | Arg | Gly | His | Ser | Thr | Pro | Ser |
| 35 |     |     | 130 |     |     |     |     | 135 |     |     |     | 140 |     |     |     |     |
|    | Pro | Tyr | Lys | Gly | Leu | Val | Leu | Ser | Lys | Leu | His | Glu | Thr | Leu | Arg | Thr |

|    |                                                                 |     |     |     |     |     |     |
|----|-----------------------------------------------------------------|-----|-----|-----|-----|-----|-----|
|    | 145                                                             |     | 150 |     | 155 |     | 160 |
|    | His Thr Arg Leu Val Ser Leu Phe Asn Thr Thr Leu Thr Arg Leu His |     |     |     |     |     |     |
|    |                                                                 | 165 |     | 170 |     | 175 |     |
| 5  | Glu Val Ser Ala Gln Asn Pro Thr Asn Cys Trp Met Cys Leu Pro Leu |     |     |     |     |     |     |
|    |                                                                 | 180 |     | 185 |     | 190 |     |
|    | His Phe Arg Pro Tyr Ile Ser Ile Pro Val Pro Glu Gln Trp Asn Asn |     |     |     |     |     |     |
|    |                                                                 | 195 |     | 200 |     | 205 |     |
|    | Phe Ser Thr Glu Ile Asn Thr Thr Ser Val Leu Val Gly Pro Leu Val |     |     |     |     |     |     |
|    |                                                                 | 210 |     | 215 |     | 220 |     |
| 10 | Ser Asn Leu Glu Ile Thr His Thr Ser Asn Leu Thr Cys Val Lys Phe |     |     |     |     |     |     |
|    |                                                                 | 225 |     | 230 |     | 235 |     |
|    | Ser Asn Thr Ile Asp Thr Thr Ser Ser Gln Cys Ile Arg Trp Val Thr |     |     |     |     |     |     |
|    |                                                                 | 245 |     | 250 |     | 255 |     |
|    | Pro Pro Thr Arg Ile Val Cys Leu Pro Ser Gly Ile Phe Phe Val Cys |     |     |     |     |     |     |
| 15 |                                                                 | 260 |     | 265 |     | 270 |     |
|    | Gly Thr Ser Ala Tyr His Cys Leu Asn Gly Ser Ser Glu Ser Met Cys |     |     |     |     |     |     |
|    |                                                                 | 275 |     | 280 |     | 285 |     |
|    | Phe Leu Ser Phe Leu Val Pro Pro Met Thr Ile Tyr Thr Glu Gln Asp |     |     |     |     |     |     |
|    |                                                                 | 290 |     | 295 |     | 300 |     |
| 20 | Leu Tyr Asn His Val Val Pro Lys Pro His Asn Lys Arg Val Pro Ile |     |     |     |     |     |     |
|    |                                                                 | 305 |     | 310 |     | 315 |     |
|    | Leu Pro Phe Val Ile Arg Ala Gly Val Leu Gly Arg Leu Gly Thr Gly |     |     |     |     |     |     |
|    |                                                                 | 325 |     | 330 |     | 335 |     |
|    | Ile Gly Ser Ile Thr Thr Ser Thr Gln Phe Tyr Tyr Lys Leu Ser Gln |     |     |     |     |     |     |
| 25 |                                                                 | 340 |     | 345 |     | 350 |     |
|    | Glu Ile Asn Gly Asp Met Glu Gln Val Thr Asp Ser Leu Val Thr Leu |     |     |     |     |     |     |
|    |                                                                 | 355 |     | 360 |     | 365 |     |
|    | Gln Asp Gln Leu Asn Ser Leu Ala Ala Val Val Leu Gln Asn Arg Arg |     |     |     |     |     |     |
|    |                                                                 | 370 |     | 375 |     | 380 |     |
| 30 | Ala Leu Asp Leu Leu Thr Ala Lys Arg Gly Gly Thr Cys Leu Phe Leu |     |     |     |     |     |     |
|    |                                                                 | 385 |     | 390 |     | 395 |     |
|    | Gly Glu Glu Arg Cys Tyr Tyr Val Asn Gln Ser Arg Ile Val Thr Glu |     |     |     |     |     |     |
|    |                                                                 | 405 |     | 410 |     | 415 |     |
|    | Lys Val Lys Glu Ile Arg Asp Arg Ile Gln Cys Arg Ala Glu Glu Leu |     |     |     |     |     |     |
| 35 |                                                                 | 420 |     | 425 |     | 430 |     |
|    | Gln Asn Thr Glu Arg Trp Gly Leu Leu Ser Gln Trp Met Pro Trp Val |     |     |     |     |     |     |

7

435                                      440                                      445  
 Leu Pro Phe Leu Gly Pro Leu Ala Ala Leu Ile Leu Leu Leu Leu Phe  
 450                                      455                                      460  
 Gly Pro Cys Ile Phe Asn Leu Leu Val Lys Phe Val Ser Ser Arg Ile  
 5                                      465                                      470                                      475                                      480  
 Glu Ala Val Lys Leu Gln Met Val Leu Gln Met Glu Pro  
                                          485                                      490

(2) INFORMATIONS POUR LA SEQ ID NO: 119:

10 (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 32 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

15 (ii) TYPE DE MOLECULE: ADNc

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 119:

TCAAAATCGA AGAGCTTTAG ACTTGCTAAC CG

32

(2) INFORMATIONS POUR LA SEQ ID NO: 120:

20 (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 1329 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

25 (ii) TYPE DE MOLECULE: ADNc

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 120:

TCAAAATCGA AGAGCTTTAG ACTTGCTAAC CGCCAAAAGA GGGGGAACCT GTTTATTTTT 60  
 AGGGGAAGAA TGCTGTTAGT ATGTTAATCA ATCTGGAATC ATTACTGAGA AAGTTAAAGA 120  
 AATTTGAGAT CGAATATAAT GTAGAGCAGA GGACCTTCAA AACACTGCAC CCTGGGGCCT 180  
 30 CCTCAGCCAA TGGATGCCCT GGA CTCTCTCCC CTTCTTAGGA CCTCTAGCAG CTATAATATT 240  
 TTTACTCCTC TTTGGACCCT GTATCTTCAA CTTCTTGTGTT AAGTTTGTCT CTTCCAGAAT 300  
 TGAAGCTGTA AAGCTACAAA TAGTTCTTCA AATGGAACCC CAGATGCAGT CCATGACTAA 360  
 AATCTACCGT GGACCCCTGG ACCGGCCTGC TAGACTATGC TCTGATGTTA ATGACATTGA 420  
 AGTCACCCCT CCCGAGGAAA TCTCAACTGC ACAACCCCTA CTACACTCCA ATTCAGTAGG 480  
 35 AAGCAGTTAG AGCAGTTGTC AGCCAACCTC CCCAACAGTA CTTGGGTTTT CCTGTTGAGA 540  
 GGGTGGACTG AGAGACAGGA CTAGCTGGAT TTCCTAGGCT GACTAAGAAT CCCNAAGCCT 600

ANCTGGGAAG GTGACCGCAT CCATCTTTAA ACATGGGGCT TGCAACTTAG CTCACACCCG 660  
 ACCAATCAGA GAGCTCACTA AAATGCTAAT CAGGCAAAAA CAGGAGGTAA AGCAATAGCC 720  
 AATCATCTAT TGCCTGAGAG CACAGCGGGA AGGACAAGGA TTGGGATATA AACTCAGGCA 780  
 TTCAAGCCAG CAACAGCAAC CCCCTTTGGG TCCCCTCCCA TTGTATGGGA GCTCTGTTTT 840  
 5 CACTCTATTT CACTCTATTA AATCATGCAA CTGCACTCTT CTGGTCCGTG TTTTTTATGG 900  
 CTCAAGCTGA GCTTTTGTTC GCCATCCACC ACTGCTGTTT GCCACCGTCA CAGACCCGCT 960  
 GCTGACTTCC ATCCCTTTGG ATCCAGCAGA GTGTCCACTG TGCTCCTGAT CCAGCGAGGT 1020  
 ACCCATTGCC ACTCCCGATC AGGCTAAAGG CTTGCCATTG TTCCTGCATG GCTAAGTGCC 1080  
 TGGGTTTGTC CTAATAGAAC TGAACACTGG TCACTGGGTT CCATGGTTCT CTTCCATGAC 1140  
 10 CCACGGCTTC TAATAGAGCT ATAACACTCA CCGCATGGCC CAAGATTCCA TTCCTTGGTA 1200  
 TCTGTGAGGC CAAGAACCCC AGGTCAGAGA ANGTGAGGCT TGCCACCATT TGGGAAGTGG 1260  
 CCCACTGCCA TTTTGGTAGC GGCCCACCAC CATCTTGGGA GCTGTGGGAG CAAGGATCCC 1320  
 CCAGTAACA 1329

15 (2) INFORMATIONS POUR LA SEQ ID NO: 121:

(i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 162 acides aminés

(B) TYPE: acide aminé

(C) NOMBRE DE BRINS: simple

20 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: peptide

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 121:

Gln Asn Arg Arg Ala Leu Asp Leu Leu Thr Ala Lys Arg Gly Gly Thr  
 1 5 10 15  
 25 Cys Leu Phe Leu Gly Glu Glu Cys Cys Xaa Tyr Val Asn Gln Ser Gly  
 20 25 30  
 Ile Ile Thr Glu Lys Val Lys Glu Ile Xaa Asp Arg Ile Xaa Cys Arg  
 35 40 45  
 Ala Glu Asp Leu Gln Asn Thr Ala Pro Trp Gly Leu Leu Ser Gln Trp  
 30 50 55 60  
 Met Pro Trp Thr Leu Pro Phe Leu Gly Pro Leu Ala Ala Ile Ile Phe  
 65 70 75 80  
 Leu Leu Leu Phe Gly Pro Cys Ile Phe Asn Phe Leu Val Lys Phe Val  
 85 90 95  
 35 Ser Ser Arg Ile Glu Ala Val Lys Leu Gln Ile Val Leu Gln Met Glu  
 100 105 110

9

Pro Gln Met Gln Ser Met Thr Lys Ile Tyr Arg Gly Pro Leu Asp Arg  
                   115                                  120                                  125  
 Pro Ala Arg Leu Cys Ser Asp Val Asn Asp Ile Glu Val Thr Pro Pro  
                   130                                  135                                  140  
 5 Glu Glu Ile Ser Thr Ala Gln Pro Leu Leu His Ser Asn Ser Val Gly  
                   145                                  150                                  155                                  160  
 Ser Ser

## (2) INFORMATIONS POUR LA SEQ ID NO: 122:

## 10 (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 21 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

## 15 (ii) TYPE DE MOLECULE: ADNc

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 122:

GGCATTGATA GCACCCATCA G

21

## (2) INFORMATIONS POUR LA SEQ ID NO: 123:

## 20 (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 21 paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

## 25 (ii) TYPE DE MOLECULE: ADNc

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 123:

CATGTCACCA GGGTGGAATA G

21

## (2) INFORMATIONS POUR LA SEQ ID NO: 124:

## 30 (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: paires de bases
- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

## 35 (ii) TYPE DE MOLECULE: ADNc

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 124:

## (2) INFORMATIONS POUR LA SEQ ID NO: 124:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- 5 (A) LONGUEUR: 758 paires de bases  
 (B) TYPE: nucléotide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADNc

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 124:

10 GGCATTGATA GCACCCATCA GATGGCCAAA TCATTATTTA CTGGACCAGG CCTTTTCAAA 60  
 ACTATCAAGC AGATAGGGCC CGTGAAGCAT GCCAAAGAAA TAATCCCCTG CCTTATCGCC 120  
 ATGTTCTTC AGGAGAACAA AGAACAGGCC ATTACCCAGG GGAAGACTGG CAACTAGATT 180  
 TTACCCACAT GGCCAAATGT CAGGGATTTC AGCATCTACT AGTCTGGGCA GATACTTTCA 240  
 CTGGTTGGGT GGAGTCTTCT CCTTGTTAGGA CAGAAAAGAC CCAAGAGGTA ATAAAGGCAC 300  
 15 TAATGAAATA ATTCCCAGAT TTGGACTTCC CCCAGGATTA CAGGGTGACA ATGGCCCCGC 360  
 TTTCAAGGCT GCAGTAACCC AGGGAGTATC CCAGGTGTTA GGCATACAAT ATCACTTACA 420  
 CTGTGCCTGG AGGCCACAAT CCTCCAGAAA AGTCAAGAAA ATGAATGAAA CACTCAAAGA 480  
 TCTAAAAAAG CTAACCCAAG AAACCCACAT TGCATGACCT GTTCTGTTGC CTATAACCTT 540  
 ACTAAGAATC CATAACTATC CCCCCAAAAG CAGGACTTAG CCCATACGAG ATGCTATATG 600  
 20 GATGGCCTTT CTAACCAAT GACCTTGTGC TTGACTGAGA AATGGCCAAC TTAGTTGCAG 660  
 ACATCACCTC CTTAGCCAAA TATCAACAAG TTCTTAAAC ATCACAGGGA ACCTGTCCCC 720  
 GAGAGGAGGG AAAGGAACTA TTCCACCCTG GTGACATG 758

## (2) INFORMATIONS POUR LA SEQ ID NO: 126:

## 25 (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 25 paires de bases  
 (B) TYPE: nucléotide  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

## 30 (ii) TYPE DE MOLECULE: ADNc

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 126:

CGGACATCCA AAGTGATGGG AAACG

25

## (2) INFORMATIONS POUR LA SEQ ID NO: 127:

## 35 (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 26 paires de bases

11

- (B) TYPE: nucléotide
- (C) NOMBRE DE BRINS: simple
- (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

5 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 127:

GGACAGGAAA GTAAGACTGA GAAGGC

26

(2) INFORMATIONS POUR LA SEQ ID NO: 128:

(i) CARACTERISTIQUES DE LA SEQUENCE:

10 (A) LONGUEUR: 26 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

15 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 128:

CCTAGAACGT ATTCTGGAGA ATTGGG

26

(2) INFORMATIONS POUR LA SEQ ID NO: 129:

(i) CARACTERISTIQUES DE LA SEQUENCE:

20 (A) LONGUEUR: 26 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

25 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 129:

TGGCTCTCAA TGGTCAAACA TACCCG

26

(2) INFORMATIONS POUR LA SEQ ID NO: 130:

(i) CARACTERISTIQUES DE LA SEQUENCE:

30 (A) LONGUEUR: 1511 paires de bases

(B) TYPE: nucléotide

(C) NOMBRE DE BRINS: simple

(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

35 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 130:

CCTAGAACGT ATTCTGGAGA ATTGGGACCA ATGTGACACT CAGACGCTAA GAAAGAAACG

60

12

```

ATTATATTC TTCTGCAGTA CCGCCTGGCC ACAATATCCT CTTCAAGGGA GAGAAACCTG 120
GCTTCCTGAG GGAAGTATAA ATTATAACAT CATCTTACAG CTAGACCTCT TCTGTAGAAA 180
GGAGGGCAAA TGGAGTGAAG TGCCATATGT GCAAACCTTC TTTTCATTAA GAGACAACTC 240
ACAATTATGT AAAAAGTGTG GTTTATGCCC TACAGGAAGC CCTCAGAGTC CACCTCCCTA 300
5 CCCAGCGTC CCCTCCCCGA CTCCTTCCTC AACTAATAAG GACCCCCCTT TAACCCAAAC 360
GGTCCAAAAG GAGATAGACA AAGGGGTAAA CAATGAACCA AAGAGTGCCA ATATTCCCCG 420
ATTATGCCCC CTCCAAGCAG TGAGAGGAGG AGAATTCGGC CCAGCCAGAG TGCCTGTACC 480
TTTTTCTCTC TCAGACTTAA AGCAAATTAA AATAGACCTA GGTAATTCT CAGATAACCC 540
TGACGGCTAT ATTGATGTTT TACAAGGGTT AGGACAATCC TTTGATCTGA CATGGAGAGA 600
10 TATAATGTTA CTAATAATC AGACACTAAC CCCAAATGAG AGAAGTGCCG CTGTAAGTGC 660
AGCCCGAGAG TTTGGCGATC TTTGGTATCT CAGTCAGGCC AACAATAGGA TGACAACAGA 720
GGAAAGAACA ACTCCCACAG GCCAGCAGGC AGTCCCAGT GTAGACCCTC ATTGGGACAC 780
AGAATCAGAA CATGGAGATT GGTGCCACAA ACATTTGCTA ACTTGCGTGC TAGAAGGACT 840
GAGGAAAAC AGGAAGAAGC CTATGAATTA CTCAATGATG TCCACTATAA CACAGGGAAA 900
15 GGAAGAAAT CTTACTGCTT TTCTGGACAG ACTAAGGGAG GCATTGAGGA AGCATACCTC 960
CCTGTACCT GACTCTATTG AAGGCCAACT AATCTTAAAG GATAAGTTTA TCACTCAGTC 1020
AGCTGCAGAC ATTAGAAAAA ACTTCAAAG TCTGCCTTAG GCCCGGAGCA GAACTTAGAA 1080
ACCCTATTTA ACTTGGCATC CTCAGTTTTT TATAATAGAG ATCAGGAGGA GCAGGCGAAA 1140
CGGGACAAAC GGGATAAAAA AAAAAGGGGG GGTCCACTAC TTTAGTCATG GCCCTCAGGC 1200
20 AAGCAGACTT TGGAGGCTCT GCAAAGGGA AAAGCTGGGC AAATCAAATG CCTAATAGGG 1260
CTGGCTTCCA GTGCGGTCTA CAAGGACACT TTAAAAAGA TTATCCAAGT AGAAATAAGC 1320
CGCCCCCTTG TCCATGCCCC TTACGTCAAG GGAATCACTG GAAGGCCAC TGCCCCAGGG 1380
GATGAAGATA CTCTGAGTCA GAAGCCATTA ACCAGATGAT CCAGCAGCAG GACTGAGGGT 1440
GCCCCGGGCG AGCGCCAGCC CATGCCATCA CCCTCACAGA GCCCCGGGTA TGTTTGACCA 1500
25 TTGAGAGCCA A 1511

```

(2) INFORMATIONS POUR LA SEQ ID NO: 131:

(i) CARACTERISTIQUES DE LA SEQUENCE:

- 30 (A) LONGUEUR: 352 acides aminés  
 (B) TYPE: acide aminé  
 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: peptide

(xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 131:

35 Leu Glu Arg Ile Leu Glu Asn Trp Asp Gln Cys Asp Thr Gln Thr Leu  
 1 5 10 15



13

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | Arg | Lys | Lys | Arg | Phe | Ile | Phe | Phe | Cys | Ser | Thr | Ala | Trp | Pro | Gln | Tyr |
|    |     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
|    | Pro | Leu | Gln | Gly | Arg | Glu | Thr | Trp | Leu | Pro | Glu | Gly | Ser | Ile | Asn | Tyr |
|    |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| 5  | Asn | Ile | Ile | Leu | Gln | Leu | Asp | Leu | Phe | Cys | Arg | Lys | Glu | Gly | Lys | Trp |
|    |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
|    | Ser | Glu | Val | Pro | Tyr | Val | Gln | Thr | Phe | Phe | Ser | Leu | Arg | Asp | Asn | Ser |
|    | 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     | 80  |     |
|    | Gln | Leu | Cys | Lys | Lys | Cys | Gly | Leu | Cys | Pro | Thr | Gly | Ser | Pro | Gln | Ser |
| 10 |     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
|    | Pro | Pro | Pro | Tyr | Pro | Ser | Val | Pro | Ser | Pro | Thr | Pro | Ser | Ser | Thr | Asn |
|    |     |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |
|    | Lys | Asp | Pro | Pro | Leu | Thr | Gln | Thr | Val | Gln | Lys | Glu | Ile | Asp | Lys | Gly |
|    |     |     |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |
| 15 | Val | Asn | Asn | Glu | Pro | Lys | Ser | Ala | Asn | Ile | Pro | Arg | Leu | Cys | Pro | Leu |
|    |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
|    | Gln | Ala | Val | Arg | Gly | Gly | Glu | Phe | Gly | Pro | Ala | Arg | Val | Pro | Val | Pro |
|    | 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     | 160 |     |
|    | Phe | Ser | Leu | Ser | Asp | Leu | Lys | Gln | Ile | Lys | Ile | Asp | Leu | Gly | Lys | Phe |
| 20 |     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
|    | Ser | Asp | Asn | Pro | Asp | Gly | Tyr | Ile | Asp | Val | Leu | Gln | Gly | Leu | Gly | Gln |
|    |     |     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |
|    | Ser | Phe | Asp | Leu | Thr | Trp | Arg | Asp | Ile | Met | Leu | Leu | Leu | Asn | Gln | Thr |
|    |     |     |     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |
| 25 | Leu | Thr | Pro | Asn | Glu | Arg | Ser | Ala | Ala | Val | Thr | Ala | Ala | Arg | Glu | Phe |
|    |     | 210 |     |     |     |     | 215 |     |     |     |     |     |     | 220 |     |     |
|    | Gly | Asp | Leu | Trp | Tyr | Leu | Ser | Gln | Ala | Asn | Asn | Arg | Met | Thr | Thr | Glu |
|    | 225 |     |     |     |     | 230 |     |     |     |     |     | 235 |     |     | 240 |     |
|    | Glu | Arg | Thr | Thr | Pro | Thr | Gly | Gln | Gln | Ala | Val | Pro | Ser | Val | Asp | Pro |
| 30 |     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
|    | His | Trp | Asp | Thr | Glu | Ser | Glu | His | Gly | Asp | Trp | Cys | His | Lys | His | Leu |
|    |     |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |
|    | Leu | Thr | Cys | Val | Leu | Glu | Gly | Leu | Arg | Lys | Thr | Arg | Lys | Lys | Pro | Met |
|    |     |     | 275 |     |     |     |     |     | 280 |     |     |     |     | 285 |     |     |
| 35 | Asn | Tyr | Ser | Met | Met | Ser | Thr | Ile | Thr | Gln | Gly | Lys | Glu | Glu | Asn | Leu |
|    |     |     |     |     | 290 |     |     |     | 295 |     |     |     |     | 300 |     |     |

14

|   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
|   | Thr | Ala | Phe | Leu | Asp | Arg | Leu | Arg | Glu | Ala | Leu | Arg | Lys | His | Thr | Ser |  |
|   | 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |  |
|   | Leu | Ser | Pro | Asp | Ser | Ile | Glu | Gly | Gln | Leu | Ile | Leu | Lys | Asp | Lys | Phe |  |
|   |     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |  |
| 5 | Ile | Thr | Gln | Ser | Ala | Ala | Asp | Ile | Arg | Lys | Asn | Phe | Lys | Ser | Leu | Pro |  |
|   |     |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |  |

(2) INFORMATION POUR LA SEQ ID NO: 132:

(i) CARACTERISTIQUES DE LA SEQUENCE:

10 (A) LONGUEUR: 30 paires de bases  
(B) TYPE: nucléotide  
(C) NOMBRE DE BRINS: simple  
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

15 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 132:

TGCTGGAATT CGGGATCCTA GAACGTATTC

30

(2) INFORMATION POUR LA SEQ ID NO: 133:

(i) CARACTERISTIQUES DE LA SEQUENCE:

20 (A) LONGUEUR: 30 paires de bases  
(B) TYPE: nucléotide  
(C) NOMBRE DE BRINS: simple  
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: ADNc

25 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 133:

AGTTCTGCTC CGAAGCTTAG GCAGACTTTT

30

(2) INFORMATION POUR LA SEQ ID NO: 135:

(i) CARACTERISTIQUES DE LA SEQUENCE:

30 (A) LONGUEUR: 398 acides aminés  
(B) TYPE: acide aminé  
(C) NOMBRE DE BRINS: simple  
(D) CONFIGURATION: linéaire

(ii) TYPE DE MOLECULE: peptide

35 (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 135:

Met Gly Ser Ser His His His His His His Ser Ser Gly Leu Val Pro

15

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | 1   |     | 5   |     | 10  |     | 15  |     |     |     |     |     |     |     |     |     |
|    | Arg | Gly | Ser | His | Met | Ala | Ser | Met | Thr | Gly | Gly | Gln | Gln | Met | Gly | Arg |
|    |     |     |     | 20  |     |     |     | 25  |     |     |     |     | 30  |     |     |     |
|    | Ile | Leu | Glu | Arg | Ile | Leu | Glu | Asn | Trp | Asp | Gln | Cys | Asp | Thr | Gln | Thr |
| 5  |     |     | 35  |     |     |     | 40  |     |     |     | 45  |     |     |     |     |     |
|    | Leu | Arg | Lys | Lys | Arg | Phe | Ile | Phe | Phe | Cys | Ser | Thr | Ala | Trp | Pro | Gln |
|    |     | 50  |     |     |     | 55  |     |     |     | 60  |     |     |     |     |     |     |
|    | Tyr | Pro | Leu | Gln | Gly | Arg | Glu | Thr | Trp | Leu | Pro | Glu | Gly | Ser | Ile | Asn |
|    | 65  |     |     |     | 70  |     |     |     | 75  |     |     |     | 80  |     |     |     |
| 10 | Tyr | Asn | Ile | Ile | Leu | Gln | Leu | Asp | Leu | Phe | Cys | Arg | Lys | Glu | Gly | Lys |
|    |     |     |     | 85  |     |     |     | 90  |     |     |     | 95  |     |     |     |     |
|    | Trp | Ser | Glu | Val | Pro | Tyr | Val | Gln | Thr | Phe | Phe | Ser | Leu | Arg | Asp | Asn |
|    |     |     | 100 |     |     |     | 105 |     |     |     |     | 110 |     |     |     |     |
|    | Ser | Gln | Leu | Cys | Lys | Lys | Cys | Gly | Leu | Cys | Pro | Thr | Gly | Ser | Pro | Gln |
| 15 |     |     | 115 |     |     |     | 120 |     |     |     |     | 125 |     |     |     |     |
|    | Ser | Pro | Pro | Pro | Tyr | Pro | Ser | Val | Pro | Ser | Pro | Thr | Pro | Ser | Ser | Thr |
|    |     | 130 |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |     |
|    | Asn | Lys | Asp | Pro | Pro | Leu | Thr | Gln | Thr | Val | Gln | Lys | Glu | Ile | Asp | Lys |
|    | 145 |     |     |     | 150 |     |     |     | 155 |     |     | 160 |     |     |     |     |
| 20 | Gly | Val | Asn | Asn | Glu | Pro | Lys | Ser | Ala | Asn | Ile | Pro | Arg | Leu | Cys | Pro |
|    |     |     |     | 165 |     |     |     | 170 |     |     |     | 175 |     |     |     |     |
|    | Leu | Gln | Ala | Val | Arg | Gly | Gly | Glu | Phe | Gly | Pro | Ala | Arg | Val | Pro | Val |
|    |     |     | 180 |     |     |     | 185 |     |     |     | 190 |     |     |     |     |     |
|    | Pro | Phe | Ser | Leu | Ser | Asp | Leu | Lys | Gln | Ile | Lys | Ile | Asp | Leu | Gly | Lys |
| 25 |     |     | 195 |     |     |     | 200 |     |     |     |     | 205 |     |     |     |     |
|    | Phe | Ser | Asp | Asn | Pro | Asp | Gly | Tyr | Ile | Asp | Val | Leu | Gln | Gly | Leu | Gly |
|    |     | 210 |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |     |
|    | Gln | Ser | Phe | Asp | Leu | Thr | Trp | Arg | Asp | Ile | Met | Leu | Leu | Leu | Asn | Gln |
|    | 225 |     |     |     | 230 |     |     |     | 235 |     |     | 240 |     |     |     |     |
| 30 | Thr | Leu | Thr | Pro | Asn | Glu | Arg | Ser | Ala | Ala | Val | Thr | Ala | Ala | Arg | Glu |
|    |     |     | 245 |     |     |     | 250 |     |     |     | 255 |     |     |     |     |     |
|    | Phe | Gly | Asp | Leu | Trp | Tyr | Leu | Ser | Gln | Ala | Asn | Asn | Arg | Met | Thr | Thr |
|    |     | 260 |     |     |     | 265 |     |     |     |     | 270 |     |     |     |     |     |
|    | Glu | Glu | Arg | Thr | Thr | Pro | Thr | Gly | Gln | Gln | Ala | Val | Pro | Ser | Val | Asp |
| 35 |     |     | 275 |     |     |     | 280 |     |     |     | 285 |     |     |     |     |     |
|    | Pro | His | Trp | Asp | Thr | Glu | Ser | Glu | His | Gly | Asp | Trp | Cys | His | Lys | His |

16

290                      295                      300  
 Leu Leu Thr Cys Val Leu Glu Gly Leu Arg Lys Thr Arg Lys Lys Pro  
 305                      310                      315                      320  
 Met Asn Tyr Ser Met Met Ser Thr Ile Thr Gln Gly Lys Glu Glu Asn  
 5                      325                      330                      335  
 Leu Thr Ala Phe Leu Asp Arg Leu Arg Glu Ala Leu Arg Lys His Thr  
 340                      345                      350  
 Ser Leu Ser Pro Asp Ser Ile Glu Gly Gln Leu Ile Leu Lys Asp Lys  
 355                      360                      365  
 10 Phe Ile Thr Gln Ser Ala Ala Asp Ile Arg Lys Asn Phe Lys Ser Leu  
 370                      375                      380  
 Pro Lys Leu Ala Ala Ala Leu Glu His His His His His His  
 385                      390                      395

## 15 (2) INFORMATIONS POUR LA SEQ ID NO: 137:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

(A) LONGUEUR: 378 acides aminés

(B) TYPE: acide aminé

(C) NOMBRE DE BRINS: simple

20 (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: peptide

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 137:

Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Ile Leu Glu Arg  
 1                      5                      10                      15  
 25 Ile Leu Glu Asn Trp Asp Gln Cys Asp Thr Gln Thr Leu Arg Lys Lys  
 20                      25                      30  
 Arg Phe Ile Phe Phe Cys Ser Thr Ala Trp Pro Gln Tyr Pro Leu Gln  
 35                      40                      45  
 Gly Arg Glu Thr Trp Leu Pro Glu Gly Ser Ile Asn Tyr Asn Ile Ile  
 30                      50                      55                      60  
 Leu Gln Leu Asp Leu Phe Cys Arg Lys Glu Gly Lys Trp Ser Glu Val  
 65                      70                      75                      80  
 Pro Tyr Val Gln Thr Phe Phe Ser Leu Arg Asp Asn Ser Gln Leu Cys  
 85                      90                      95  
 35 Lys Lys Cys Gly Leu Cys Pro Thr Gly Ser Pro Gln Ser Pro Pro Pro  
 100                      105                      110

|    | Tyr | Pro | Ser | Val | Pro | Ser | Pro | Thr | Pro | Ser | Ser | Thr | Asn | Lys | Asp | Pro |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    |     | 115 |     |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
|    | Pro | Leu | Thr | Gln | Thr | Val | Gln | Lys | Glu | Ile | Asp | Lys | Gly | Val | Asn | Asn |
|    |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| 5  | Glu | Pro | Lys | Ser | Ala | Asn | Ile | Pro | Arg | Leu | Cys | Pro | Leu | Gln | Ala | Val |
|    | 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
|    | Arg | Gly | Gly | Glu | Phe | Gly | Pro | Ala | Arg | Val | Pro | Val | Pro | Phe | Ser | Leu |
|    |     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
|    | Ser | Asp | Leu | Lys | Gln | Ile | Lys | Ile | Asp | Leu | Gly | Lys | Phe | Ser | Asp | Asn |
| 10 |     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
|    | Pro | Asp | Gly | Tyr | Ile | Asp | Val | Leu | Gln | Gly | Leu | Gly | Gln | Ser | Phe | Asp |
|    |     | 195 |     |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
|    | Leu | Thr | Trp | Arg | Asp | Ile | Met | Leu | Leu | Leu | Asn | Gln | Thr | Leu | Thr | Pro |
|    |     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| 15 | Asn | Glu | Arg | Ser | Ala | Ala | Val | Thr | Ala | Ala | Arg | Glu | Phe | Gly | Asp | Leu |
|    | 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
|    | Trp | Tyr | Leu | Ser | Gln | Ala | Asn | Asn | Arg | Met | Thr | Thr | Glu | Glu | Arg | Thr |
|    |     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
|    | Thr | Pro | Thr | Gly | Gln | Gln | Ala | Val | Pro | Ser | Val | Asp | Pro | His | Trp | Asp |
| 20 |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
|    | Thr | Glu | Ser | Glu | His | Gly | Asp | Trp | Cys | His | Lys | His | Leu | Leu | Thr | Cys |
|    |     | 275 |     |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
|    | Val | Leu | Glu | Gly | Leu | Arg | Lys | Thr | Arg | Lys | Lys | Pro | Met | Asn | Tyr | Ser |
|    |     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| 25 | Met | Met | Ser | Thr | Ile | Thr | Gln | Gly | Lys | Glu | Glu | Asn | Leu | Thr | Ala | Phe |
|    | 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
|    | Leu | Asp | Arg | Leu | Arg | Glu | Ala | Leu | Arg | Lys | His | Thr | Ser | Leu | Ser | Pro |
|    |     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |
|    | Asp | Ser | Ile | Glu | Gly | Gln | Leu | Ile | Leu | Lys | Asp | Lys | Phe | Ile | Thr | Gln |
| 30 |     |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |
|    | Ser | Ala | Ala | Asp | Ile | Arg | Lys | Asn | Phe | Lys | Ser | Leu | Pro | Lys | Leu | Ala |
|    |     | 355 |     |     |     |     |     | 360 |     |     |     |     | 365 |     |     |     |
|    | Ala | Ala | Leu | Glu | His | His | His | His | His | His | His |     |     |     |     |     |
|    |     | 370 |     |     |     |     |     | 375 |     |     |     |     |     |     |     |     |

## FEUILLE DE REMPLACEMENT (REGLE 26)

18

- (i) CARACTERISTIQUES DE LA SEQUENCE:  
    (A) LONGUEUR: 25 paires de bases  
    (B) TYPE: nucléotide  
    (C) NOMBRE DE BRINS: simple  
5      (D) CONFIGURATION: linéaire  
    (ii) TYPE DE MOLECULE: ADNc  
    (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 138:  
CTTGGAGGGT GCATAACCAG GGAAT 25
- 10 (2) INFORMATIONS POUR LA SEQ ID NO: 139:  
    (i) CARACTERISTIQUES DE LA SEQUENCE:  
        (A) LONGUEUR: 20 paires de bases  
        (B) TYPE: nucléotide  
        (C) NOMBRE DE BRINS: simple  
15      (D) CONFIGURATION: linéaire  
    (ii) TYPE DE MOLECULE: ADNc  
    (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 139:  
TGTCCGCTGT GTCCTGATC 20
- 20 (2) INFORMATIONS POUR LA SEQ ID NO: 140:  
    (i) CARACTERISTIQUES DE LA SEQUENCE:  
        (A) LONGUEUR: 25 paires de bases  
        (B) TYPE: nucléotide  
        (C) NOMBRE DE BRINS: simple  
25      (D) CONFIGURATION: linéaire  
    (ii) TYPE DE MOLECULE: ADNc  
    (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 140:  
CTATGTCCTT TTGGACTGTT TGGGT 25
- 30 (2) INFORMATIONS POUR LA SEQ ID NO: 141:  
    (i) CARACTERISTIQUES DE LA SEQUENCE:  
        (A) LONGUEUR: 764 paires de bases  
        (B) TYPE: nucléotide  
        (C) NOMBRE DE BRINS: simple  
35      (D) CONFIGURATION: linéaire  
    (ii) TYPE DE MOLECULE: ADNc

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 141:

TGTCCGCTGT GCTCCTGATC CAGCACAGGC GCCCATTGCC TCTCCCAATT GGGCTAAAGG 60  
 CTTGCCATTG TTCCTGCACA GCTAAGTGCC TGGGTTTCATC CTAATCGAGC TGAACACTAG 120  
 TCACTGGGTT CCACGGTTCT CTTCCATGAC CCATGGCTTC TAATAGAGCT ATAACACTCA 180  
 5 CTGCATGGTC CAAGATTCCA TTCCTTGGA TCCGTGAGAC CAAGAACCCC AGGTCAGAGA 240  
 ACACAAGGCT TGCCACCATG TTGGAAGCAG CCCACCACCA TTTTGAAGC AGCCCGCCAC 300  
 TATCTTGGGA GCTCTGGGAG CAAGGACCCC AGGTAACAAT TTGGTGACCA CGAAGGGACC 360  
 TGAATCCGCA ACCATGAAGG GATCTCCAAA GCAATTGGAA ATGTTCTCCTCC CAAGGCAAAA 420  
 ATGCCCTTAA GATGTATTCT GGAGAATTGG GACCAATTTG ACCCTCAGAC AGTAAGAAAA 480  
 10 AAATGACTTA TATTCTTCTG CAGTACCGCC CTGGCCACGA TATCCTCTTC AAGGGGGAGA 540  
 AACCTGGCCT CCTGAGGGAA GTATAAATTA TAACACCATC TTACAGCTAG ACCTGTTTTG 600  
 TAGAAAAGGA GGCAAATGGA GTGAAGTGCC ATATTTACAA ACTTTCTTTT CATTAAAGA 660  
 CAACTCGCAA TTATGTTAAC AGTGTGATTT GTGTTCTTAC ACGGAAGCCC TCAGATTCTA 720  
 CTCCCCACCC CCGGCATCTC CCCTGAATCC CTCCCCAACT TATT 764  
 15

## (2) INFORMATIONS POUR LA SEQ ID NO: 142:

## (i) CARACTERISTIQUES DE LA SEQUENCE:

- (A) LONGUEUR: 800 paires de bases  
 (B) TYPE: nucléotide  
 20 (C) NOMBRE DE BRINS: simple  
 (D) CONFIGURATION: linéaire

## (ii) TYPE DE MOLECULE: ADNc

## (xi) DESCRIPTION DE LA SEQUENCE: SEQ ID NO: 142:

TGTCCGCTGT GCTCCTGATC CAGCACAGGC GCCCATTGCC TCTCCCAATT GGGCTAAAGG 60  
 25 CTTGCCATTG TTCCTGCACA GCTAAGTGCC TGGGTTTCATC CTAATCGAGC TGAACACTAG 120  
 TCACTGGGTT CCACGGTTCT CTTCCATGAC CCATGGCTTC TAATAGAGCT ATAACACTCA 180  
 CTGCATGGTC CAAGATTCCA TTCCTTGGA TCCGTGAGAC CAAGAACCCC AGGTCAGAGA 240  
 ACACAAGGCT TGCCACCATG TTGGAAGCAG CCCACCACCA TTTTGAAGC GGCCCGCCAC 300  
 TATCTTGGGA GCTCTGGGAG CAAGGACCCC CAGGTAACAA TTTGGTGACC ACGAAGGGAC 360  
 30 CTGAATCCGC AACCATGAAG GGATCTCCAA AGCAATTGGA AATGTTCTCT CCAAGGCAAA 420  
 AATGCCCTTA AGATGTATTC TGGAGAATTG GGACCAATCT GACCCTCAGA CAGTAAGAAA 480  
 AAAAATGACT TATATTCTTC TGCAGTACCG CCTGGCCACG GATATCCTCT TCAAGGGGGA 540  
 GAAACCTGGC CTCCTGAGGG AAGTATAAAT TATAACACCA TCTTACAGCT AGACCTGTTT 600  
 TGTAGAAAAG GAGGCAAATG GAGTGAAGTG CCATATTTAC AAACCTTCTT TTCATTAAAA 660  
 35 GACAACTCGC AATTATGTAA ACAGTGTGAT TTGTGTCCTA CAGGAAGCCC TCAGATCTAC 720  
 CTCCCTACCC CGGCATCTCC CTGACTCCTT CCCCAACTAA TAAGGACCCA CTTCAGCCCA 780

20

AACAGTCCAA AAGGACATAG

800